

FORECASTING FOR NONLINEAR AND
NONSTATIONARY SYSTEMS USING INTRINSIC
FUNCTIONAL DECOMPOSITION MODELS

By

AKKARAPOL SA-NGASOONGSONG

Bachelor of Engineering in Electrical Engineering
Mahidol University
Nakhon Pathom, Thailand
2003

Master of Science in Industrial Engineering
Oklahoma State University
Stillwater, Oklahoma
2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2014

COPYRIGHT ©
By
AKKARAPOL SA-NGASOONGSONG
December, 2014

FORECASTING FOR NONLINEAR AND
NONSTATIONARY SYSTEMS USING INTRINSIC
FUNCTIONAL DECOMPOSITION MODELS

Dissertation Approved:

Dr. Satish T.S. Bukkapatnam

Dissertation Adviser

Dr. Tieming Liu

Committee Chair

Dr. Camille Deyong

Dr. William J. Kolarik

Dr. Jaebeom Kim

Dedicated to my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deepest appreciation to my academic adviser, Dr. Satish T.S. Bukkapatnam, for his patience, enthusiasm, motivation and continual support for my Ph.D. study. It has been a great honor for me to be one of his Ph.D. students. Dr. Satish has not been only a perfect mentor during the course of my Ph.D. program, but also a role model for my future career.

I wish to express infinite gratitude to my dissertation committee members, Dr. Tieming Liu, Dr. Jaebeom Kim, Dr. Camille Deyong and Dr. William J. Kolarik for their support, guidance and helpful suggestions throughout my Ph.D. study. Special thanks are also due to Dr. James Kong for his advice and continual support on my research. I feel fortunate to be able to receive guidance and support from these great and knowledgeable individuals.

Many thanks are due to all of my friends and colleagues, especially Woranat Wongdhamma, Trung Q. Le, Changqing Cheng, Rao Prahalad and Hoang Tran for their great friendship and encouragement throughout my Ph.D. study at Oklahoma State University.

My Ph.D. dissertation research was partially sponsored by General Motors and the National Science Foundation. These financial supporters are greatly appreciated.

My extreme appreciation goes to my lovely wife, Rawissada Sukittiworakul for her love, understanding, encouragement and support. She is just the most wonderful person in the world. I am grateful to have her in my life.

Finally, words cannot express the love I have for my parents, Prapath Sa-ngasoongsong and Sasithon Sa-ngasoongsong. I wish I could show them how much I appreciate them for their encouragement, sacrifice, patience and never-ending love. I dedicate my dissertation to my beloved parents.

Name: AKKARAPOL SA-NGASOONGSONG

Date of Degree: DECEMBER 2014

Title of Study: FORECASTING FOR NONLINEAR AND NONSTATIONARY
SYSTEMS USING INTRINSIC FUNCTIONAL DECOMPOSITION
MODELS

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract: The purpose of this study is to develop nonlinear and nonstationary time series forecasting methods to address modeling and prediction of real-world, complex systems. Particular emphasis has been placed on nonlinear and nonstationary time series forecasting in systems and processes that are of interest to IE researchers. Two new advanced prediction methods are developed using nonlinear decomposition techniques and a battery of advanced statistical methods. The research methodologies include empirical mode decomposition (EMD)-based prediction, structural relationship identification (SRI) methodology, and intrinsic time-scale decomposition (ITD)-based prediction. The advantages of using these prediction methods are local characteristic time scales and the use of an adaptive basis that does not require a parametric functional form (during the decomposition process). The utilization of SRI methodology in ITD-based prediction also provides a relationship identification advantage that can be used to capture the interrelationships of variables in the system for prediction application. The empirical results of using these new prediction methods have shown a significant improvement in the accuracy for customer willingness-to-pay and automobile demand prediction applications.

Contributions:

- New prediction method for univariate, nonlinear, and nonstationary time series based on empirical mode decomposition (EMD) technique. This method addresses the challenge (edge artifacts limits) of using EMD for prediction applications.
- New approach to identify relationship of multivariate, nonlinear and nonstationary time series using structural relationship identification (SRI) methodology. This approach addresses the relationship identification problem of using advanced models, including VAR and VECM in practice.
- Developed a new prediction model for multivariate, nonlinear and nonstationary time series based on SRI methodology and intrinsic time-scale decomposition (ITD).

TABLE OF CONTENTS

CHAPTER I	1
INTRODUCTION.....	1
1.1 Research Motivation	1
1.2 Research Objectives	2
1.3 Major Contributions	2
1.4 Organization of the Dissertation	3
 CHAPTER II.....	 5
BACKGROUND AND LITERATURE REVIEW.....	5
2.1 Nonlinear and Nonstationary Time Series	5
2.2 Parametric Models	8
2.2.1 Classical Autoregressive Models.....	8
2.2.2 Neural Networks	9
2.2.3 Support Vector Machine Models	10
2.2.4 Hidden or State-observer Markov Models.....	11
2.3 Nonparametric Models	11
2.3.1 State Space Neighborhood and Local Topology-based Models.....	12
2.3.2 Bayesian Nonparametric Models.....	13
2.3.3 Functional Decomposition Models.....	14
 CHAPTER III	 16
RESEARCH METHODOLOGY.....	16
3.1 Empirical Mode Decomposition (EMD) based Prediction Methodology	16
3.1.1 Empirical Mode Decomposition.....	17
3.1.2 Two-step Approach for EMD-based Prediction	20
3.2 Structural Relationship Identification (SRI) Methodology	25
3.2.1 VAR and VECM	27
3.2.2 Unit Root Tests.....	30

3.2.3	Weak Exogeneity Test.....	32
3.2.4	Cointegration Test	33
3.2.5	Granger-causality Test.....	34
3.3	Intrinsic Time-Scale Decomposition (ITD) based Prediction Methodology	35
3.3.1	Intrinsic Time-Scale Decomposition	36
3.3.2	ITD-based Prediction Methodology	38
CHAPTER IV		40
WILLINGNESS-TO-PAY PREDICTION BASED ON EMPIRICAL MODE DECOMPOSITION.....		40
4.1	Introduction.....	41
4.2	Implementation Details and Results	43
4.3	Conclusions and Suggested Future Work	45
CHAPTER V		47
MULTI-STEP AUTOMOTIVE SALES FORECASTING BASED ON STRUCTURAL RELATIONSHIP IDENTIFICATION		47
5.1	Introduction.....	48
5.2	Data.....	52
5.3	Implementation Details and Results	55
5.3.1	Unit Root.....	56
5.3.2	Weak Exogeneity	58
5.3.3	Cointegration Rank Test.....	60
5.3.4	Granger Causality Test.....	61
5.3.5	Out-of-Sample Forecasting Performance Comparison	62
5.3.6	Impulse Response Function.....	64
5.4	Conclusions and Suggested Future Work	65
CHAPTER VI		67
LONG-TERM AUTOMOBILE DEMAND PREDICTION USING INTRINSIC TIME- SCALE DECOMPOSITION.....		67
6.1	Introduction.....	68
6.2	Implementation Details and Empirical Results.....	70
6.2.1	Intrinsic Time-Scale Decomposition Step	71
6.2.2	Variable Selection Step	72

6.2.3	Prediction Step	75
6.3	Conclusions and Suggested Future Work	77
CHAPTER VII		80
COMPARISON OF NONLINEAR AND NONSTATIONARY FORECASTING MODELS		80
7.1	Comparison of Nonlinear and Nonstationary Forecasting Models	80
7.2	Case Study I: Real-time Throughput Forecasting in an Automotive Assembly Line ...	81
7.3	Case Study II: End Point Detection in Chemical Mechanical Planarization Processes.	82
CHAPTER VIII		84
CONCLUSIONS AND FUTURE WORK.....		84
8.1	Conclusions	84
8.2	Future Work	85
REFERENCES.....		88
APPENDIX.....		103

LIST OF TABLES

Table 4.1: Prediction Comparison (the first step of the two-step approach).....	43
Table 4.2: Prediction Comparison (the second step of the two-step algorithm approach)	44
Table 4.3: Comparison of prediction accuracy with the validation data set	45
Table 5.1: Summary of variables	55
Table 5.2: ADF Unit Root Test of Original and Differenced Variables	57
Table 5.3: Weak Exogeneity Tests (all variables)	59
Table 5.4: Weak Exogeneity Tests (Re-test(1))	59
Table 5.5: Weak Exogeneity Tests (Re-test(2))	59
Table 5.6: Johansen's Cointegration Rank Tests.....	60
Table 5.7: Long-Run parameter (β) and adjustment coefficient (α)	60
Table 5.8: Granger Causality Tests	61
Table 5.9: 12-step-ahead Forecasting Comparison (with trend, ADL and ARIMA(X) models) ..	62
Table 6.1: Summary of variables	71
Table 6.2: Cointegration Rank Tests (Automobile Demand and Unemployment Rate)	71
Table 6.3: ADF Unit Root Test of Original and First Differenced Variables	73
Table 6.4: Weak Exogeneity Tests (Selected Variables).....	74
Table 6.5: Weak Exogeneity Tests (Re-estimate)	74
Table 6.6: Cointegration Rank Tests (Automobile Demand and selected ITD components of Unemployment rate)	75
Table 6.7: Long-Run Parameter Beta Estimates of Cointegrating Vector (AD, H2, H3 and H4) .	75
Table 6.8: Out-of-Sample Forecasting Comparison	77

Table 7.1: Summary of the Performance of Different Prediction Models in Prior Empirical Studies.....	81
Table 7.2: Comparison of Forecasting Horizon and Accuracy of Automotive Assembly Line Throughput.....	81
Table 7.3: Comparison of Forecasting Horizon and Accuracies for the First Principal Feature and MRR.....	83

LIST OF FIGURES

Figure 2.1: Classification of nonlinear and nonstationary time series forecasting models	7
Figure 3.1: Customer WTP for a fuel economy attribute	19
Figure 3.2: (a) Amplitude Envelope and (b) Instantaneous Phase of IMF1	20
Figure 3.3: Amplitude Envelope of IMFs 2-9	20
Figure 3.4: The First Step Procedure.....	21
Figure 3.5: Detrending operation by observing the change point. (a) Standardized mean for each IMF order, (b) Trend and original series (WTP of fuel economy attribute), (c) Detrended series from detrending operation.....	22
Figure 3.6: (a) Original trend IMFs and (b) New IMFs from the first step procedure	24
Figure 3.7: Automobile Sales Modeling Framework (SRI Methodology)	27
Figure 3.8: ITD-based Prediction Methodology Framework.....	39
Figure 4.1: (a) Typical trend of WTP of a fuel economy attribute (b) Autocorrelation of WTP attribute	42
Figure 5.1: Sales of Large- and Small-Vehicle Segments	52
Figure 5.2: Economic Indicators: (a) Consumer Price Index (CPI) , (b) Unemployment Rate, (c) Gas Prices and (d) Housing Starts	53
Figure 5.3: 12-step-ahead Forecasting Comparison (2010).....	63
Figure 5.4: Model Selection Criteria Comparison (a) Average RMSE Comparison (b) Average MAPE Comparison.....	63

Figure 5.5: Impulse Response Functions from VECM: (a) Response of Y_1 from X_2 (b) Response of Y_1 from Y_2 (c) Response of Y_1 from X_4 and (d) Response of Y_1 from Y_1	65
Figure 6.1: (a) Automobile Demand and (b) Unemployment rate	70
Figure 6.2: ITD components of Unemployment Rate	72
Figure 6.3: (a) Out-of-Sample Forecasting Comparison using RMSE and (b) Out-of-Sample Forecasting Comparison using MAPE	77

CHAPTER I

INTRODUCTION

In this first chapter, the motivation for selecting this research topic is presented, followed by the objectives and major contributions of this research. The organization of this dissertation is provided at the end of this chapter.

1.1 Research Motivation

Forecasting the evolution of complex systems has been noted as one of the ten “grand challenges” of modern science [1]. Time series data from complex systems capture the dynamic behaviors and causalities of the underlying processes and provide tractable means to predict and monitor system state evolution. However, forecasting methods reported in the literature focus mostly on forecasting linear and stationary processes. Effective forecasting of future states of a complex time series system remains a challenge, mainly due to diverse combinations of the nonlinear and nonstationary dynamic behaviors exhibited by these systems.

Recently, functional decomposition models [2, 3] for nonlinear and nonstationary time series forecasting have been receiving attention in the literature. The advantages of this type of model include a complete representation of the dynamics of nonlinear and nonstationary systems based on the observed data, local characteristic time scales and the use of an adaptive basis that does not require a parametric functional form. These models do not impose any structural assumptions, and also simplify modeling efforts. In addition, they can capture drifts and nonlinear modes of

any nonlinear and nonstationary process. Functional decomposition often suppresses or disperses the effects of nonstationarity (e.g., trend and seasonality) over different functional components in the time series. Consequently, the modeling and forecasting accuracy for nonlinear and nonstationary time series is improved. This type of forecasting method offers an unprecedented opportunity to develop a prediction specifically for nonlinear and nonstationary time series systems.

1.2 Research Objectives

The proposed research addresses modeling and prediction of nonlinear and nonstationary time series in certain real-world data applications. Particular emphasis has been placed on nonlinear and nonstationary time series forecasting in systems and processes that are of interest to IE researchers. The main research objectives are as follows:

- Develop an effective prediction model to forecast real-world nonlinear and nonstationary time series over extended time horizons.
- Characterize the components of univariate, nonlinear, and nonstationary time series and develop technique to improve prediction accuracy for short- and long-term predictions.
- Develop an approach to identify and quantify the structural relationship of multivariate, nonlinear and nonstationary time series.
- Develop an effective prediction model to forecast multivariate, nonlinear, and nonstationary time series using an identified structural relationship.

1.3 Major Contributions

In this research, two new nonlinear and nonstationary prediction methodologies based on nonlinear decomposition techniques are presented. The prediction approach based on the empirical mode decomposition (EMD) technique was used to forecast univariate customer

preferences for automobile products over extended time horizons. A key aspect of this approach is to use the linear phase property of the Hilbert-Huang transform (HHT) to address the presence of edge artifact limits of the EMD method. The other methodology, specifically for multivariate, nonlinear and nonstationary time series, is to apply the intrinsic time-scale decomposition (ITD) technique to forecast automobile demand using components of selected economic indicators.

The ITD-based prediction approach has been applied to determining long-run equilibrium relationships between automobile demand and ITD components of nonlinear economic indicators. The key aspect of this new prediction approach is to use a structural relationship identification (SRI) methodology to identify a causal and long-run equilibrium relationship between a nonlinear and nonstationary time series and ITD components of related indicators and then use this identified relationship for prediction. The key contributions of this research are as follows:

- A new prediction model for univariate, nonlinear and nonstationary time series based on nonlinear decomposition technique (EMD).
- A new approach to identifying the structural relationships of multivariate, nonlinear, and nonstationary time series using a battery of advanced statistical techniques.
- A new prediction model for multivariate, nonlinear, and nonstationary time series based on SRI methodology and a nonlinear decomposition technique (ITD).
- Development of a technique to address the presence of edge artifact limits of the EMD method.

1.4 Organization of the Dissertation

The organization of the remainder of this dissertation is as follows:

Chapter II: Background and Literature Review presents an overview of nonlinear and nonstationary time series. In addition, various methods for nonlinear and nonstationary time series forecasting reported in the literature are reviewed.

Chapter III: Research Methodology presents the research methodologies used in this study. The overall research methodology consists of three main parts: EMD-based prediction, SRI methodology, and ITD-based prediction.

Chapter IV: Willingness-To-Pay Prediction based on Empirical Mode Decomposition presents implementation details and the results of using EMD-based prediction methodology, described in Chapter III, to predict customer willingness-to-pay (WTP) data for an attribute of automobiles.

Chapter V: Multi-Step Sales Forecasting in Automotive Study based on Structural Relationship Identification presents implementation details and results of using SRI methodology, described in Chapter III, to forecast automobile sales.

Chapter VI: Long-term Automobile Demand Prediction using Intrinsic Time-Scale Decomposition presents the implementation details and results of using the ITD-based prediction methodology described in Chapter III to forecast automobile demand.

Chapter VII: Comparison of Nonlinear and Nonstationary Forecasting Models presents a comparison of EMD-based prediction and other nonlinear and nonstationary forecasting models in the literature.

Chapter VIII: Conclusions and Future Work presents research contributions, general conclusions, and suggestions for future work in this area.

CHAPTER II

BACKGROUND AND LITERATURE REVIEW

This chapter begins with an overview of nonlinear and nonstationary time series, followed by a review of various methods for nonlinear and nonstationary time series forecasting reported in the literature.

2.1 Nonlinear and Nonstationary Time Series

Although real-world systems exhibit mostly nonlinear and nonstationary behaviors, the majority of forecasting methods reported in the literature assume the linearity and/or stationarity of the underlying dynamics [4], or consider simple forms of nonstationarity such as well-defined trends and variations in the first and second moments or simple forms of piecewise stationary regimes [5]. This section provides an overview and the relevant background of nonlinear and nonstationary time series.

A nonlinear time series $y(t)$ is a signal emanating from a nonlinear dynamic process. In other words, it is a partial solution of a nonlinear stochastic differential (or difference) equation of the following form

$$dx/dt = F(x, \theta, \epsilon) \quad (2.1)$$

This equation governs the evolution of process states $x(t)$ from an initial condition $x(0)$, where θ is the process parameter vector and ϵ is the system noise. The solution of Eq. (2.1) is given by

$x(t) = \phi(x(0), t)$. Here, $\phi(\cdot, \cdot)$ is called the flow, or the state transition function. Many real-world systems exhibit such nonlinear stochastic dynamics, and the solutions of such systems, referred to as nonlinear time series, exhibit non-Gaussian, multi-modality, time irreversibility and other properties [6].

Most real-world nonlinear dynamic systems operate under transient (i.e., nonstationary) conditions. From the statistical perspective, the stationarity of time series $y(t)$ requires that the joint distribution of every collection $[y(t + \tau_1), (t + \tau_2), \dots, y(t + \tau_k)]$ be invariant to $\tau_i (i = 1, 2, \dots, k)$ for any k . Even under nonstationary conditions, complex system dynamics may be treated as a concatenation of much simpler piecewise transient or near-stationary behaviors. Most commonly, nonstationarity is attributed to specific deterministic and stochastic trends in the moments. For example, a first-order nonstationary time series may be expressed as $y(t) = \mu_t + w(t)$, where μ_t is a deterministic function of time and $w(t) \sim N(\mu_w, \sigma_w^2)$ is a stationary disturbance process. The stationarity is violated, as $E[y(t)] = \mu_t$ depends on time t . According to the explicit form of the drift μ_t , the time series may exhibit growth (an upward trend) and/or decay (a downward trend). A repetitive up-and-down trend is called a cycle, and a cyclic trend that occurs at fixed and known periods is called seasonality. Additionally, processes embodying a stochastic trend such as the first-order random walk $y(t) = \alpha + y(t - 1) + w(t)$ have time-varying conditional means and variances with stationary increments. Autoregressive integrated moving average (ARIMA) models are often used to represent these patterns. The evolution of asset prices and other derivatives of financial markets is often treated as a process with time-varying and autocorrelated σ_w^2 , also known as varying volatility or heteroskedasticity processes. Autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) models are often employed to capture these behaviors.

Additional second-order nonstationary processes include *periodic time series* where there is a $T > 0$ such that the expected values and covariate satisfy the following conditions: $\mu_t = \mu_{t+T}$ for

all t and $cov(t_1, t_2) = cov(t_1 + T, t_2) = cov(t_1, t_2 + T)$ for all t_1, t_2 , and *cyclo-stationary time series* where there is a $T > 0$ such that $\mu_t = \mu_{t+T}$ for all t and $cov(t_1, t_2) = cov(t_1 + T, t_2 + T)$ for all t_1, t_2 [7]. Cyclo-stationary time series analysis is often seen in the modeling of weather parameters, such as temperatures that can be highly nonstationary during a 12-month period in a certain region, but variation in each winter for the past years can be regarded as stationary [8]. In addition to these examples, forecasting of time series with varying higher-order moments (e.g., kurtosis and skewness) has also been investigated [9].

In the following section, various methods for nonlinear and nonstationary time series forecasting reported in the literature are reviewed and categorized based on how they have been applied for forecasting real-world time series data (see Figure 2.1).

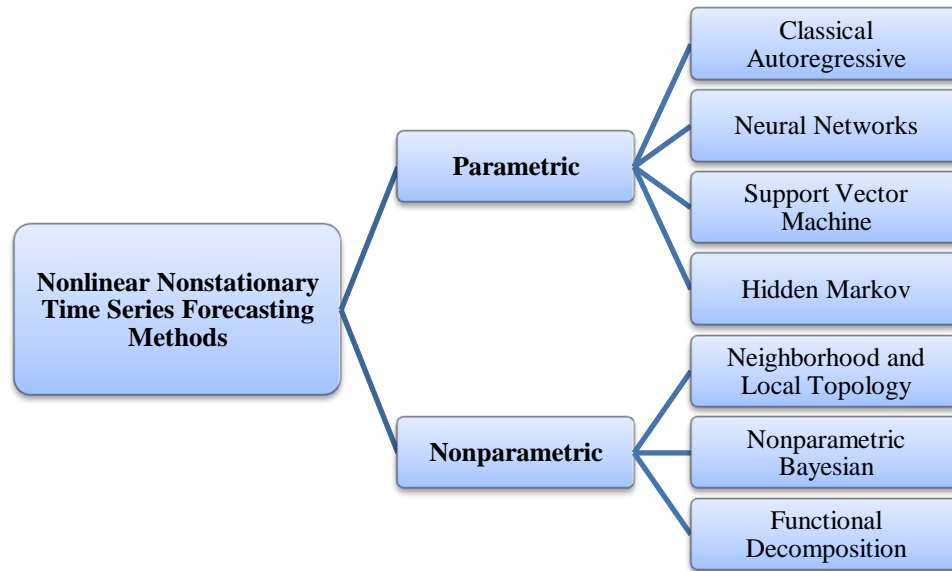


Figure 2.1: Classification of nonlinear and nonstationary time series forecasting models

Particular emphasis has been placed on nonlinear and nonstationary time series forecasting in systems and processes that are of interest to IE researchers. Broadly speaking, these forecasting methods may be classified based on the premises or the approaches to treating nonstationarity under nonlinear conditions, in that they assume either (i) a known form of the trend in the first

few moments, (ii) piecewise stationarity of the signals, (iii) progressively varying parameters, or (iv) decomposability of the signal into stationary segments in a transformed domain, and they are either parametric or nonparametric depending on whether the predictor takes a predetermined form or is constructed purely according to the data (e.g., the number of latent variables is allowed to vary).

2.2 Parametric Models

A parametric forecasting model specifies an explicit function form with a finite number of parameters θ to describe the relationship between the input, consisting of the intrinsic and exogenous variables and their autoregressive (lag) terms, and the output, consisting of the future values of the intrinsic variable $y(t + 1)$. The model parameters are estimated from the time series realizations.

2.2.1 Classical Autoregressive Models

In this group of parametric models, classical autoregressive models, such as autoregressive (AR) or autoregressive moving average (ARMA), are the most widely studied because of their flexibility in modeling many stationary processes, but they generally fail to accurately predict the evolution of nonlinear and nonstationary processes. Methods such as autoregressive integrated moving average (ARIMA), which are based on the evolution of the increment $\Delta y_t = y_{t+1} - y_t$ or $\Delta^2_{y_t}$, are used at times to remove/reduce first-order (moment) nonstationarity. However, differencing generally amplifies high frequency noise in the time series, and great effort is required to determine the order of an ARMA model. In order to incorporate nonlinearity as part of an ARMA structure, advanced models such as threshold AR models (TAR) [10, 11], self-excited threshold AR models (SETAR) [12], and smooth transition AR (STAR) [13] have been developed for nonlinear forecasting. However, these methods tend to be limited for nonlinear

stationary time series forecasting by the local linearity assumption implicit with an AR-type structure.

Multivariate time-series-based classical regression forecasting models, such as vector AR (VAR), have also been studied to consider the effects of exogenous variables. However, the nonstationary evolution of such variables, which can deviate permanently from previous states, presents additional challenges to the long-term forecasting of the target variables. A vector error correction model (VECM) can be useful in capturing such behaviors. Especially when variables are cointegrated, VAR is not a suitable model for forecasting analysis because VAR does not represent cointegrating relations for a system of variables. A special characteristic and advantage of using VECM is that the long-run equilibrium relationships of target variables can be identified and determined by the cointegration vector. The speed at which target variables return to equilibrium after a change in exogenous variables can also be estimated using VECM. The only restriction of this model for nonstationary forecasting is that target variables must be cointegrated; in other words, the cointegration vector must exist for a system of variables. For example, Sa-ngasoongsong *et al.* [14] combined VAR and VECM to forecast the long-term variation of automobile sales by estimating the cointegration vectors of the intrinsic variables.

2.2.2 Neural Networks

Neural networks (NNs) have been used for nonlinear time series forecasting in many applications [15-17]. These models do not require prior assumptions on the form of nonlinearity and are universal approximators [18], i.e., they can approximate any continuous function to an arbitrary precision. A recent review of NN models for time series forecasting has been provided by Zhang [19].

Feed-forward neural network models (FNNs) parameterized with a back-propagation algorithm have been employed for nonlinear time series forecasting [20]. They are known to outperform

traditional statistical methods such as regression and Box-Jenkins approaches in functional approximation, but they assume that the dynamics underlying time series are time-invariant. FNNs with recurrent feedback connections have also been attempted for time series forecasting [21]. Such dynamic recurrent NN (RNN) models allow forecasting of nonlinear time series occurring in various fields [22, 23]. Rao *et al.* [24] studied recurrent predictor NN (RPNN) models for one-step-ahead prediction of the nonlinear signal patterns during ultra-precision machining processes and combined particle filtering models for detection of changes in NN weights. Menezes *et al.* [25] built a recurrent network structure of nonlinear AR models with exogenous input (NARX) for multi-step forecasting of chaotic time series.

2.2.3 Support Vector Machine Models

Another class of parametric models for nonlinear and nonstationary time series forecasting is a group of support vector machine (SVM) models. SVM-based forecasting methods use a class of generalized regression models, such as support vector regression (SVR) and least-squares SVM (LS-SVM) [26], that are parameterized using convex quadratic programming methods [27]. An SVM maps the inputs x_i , which may consist of autoregressive terms of intrinsic and exogenous variables, into a higher dimensional feature space $\varphi(x_i)$. We need not compute the transform $\varphi(x_i)$ explicitly; instead we only need to estimate the inner product of the mapped patterns $k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$, where $\langle \cdot \rangle$ denotes the inner product. The inner product is expressed as a linear combination of specified kernel functions, based on which SVMs are categorized into linear, Gaussian or radial basis function (RBF), polynomial, and multilayer perceptron classifiers. A linear regressor is then constructed by minimizing the structural risk minimization (the upper bound of the generalization error), leading to better generalization than with conventional techniques [28].

Mukherjee *et al.* [29] investigated the application of SVMs for chaotic time series forecasting. They showed that SVMs have higher forecasting accuracy than NN models and employ fewer parameters. Lau and Wu [30] reviewed least squares- (LS) and RBF-based predictors and designed a local SVM (defined in the reconstructed state space) for chaotic time series forecasting. Their investigations suggest that local SVM models can provide higher accuracy for long term forecasting compared to local LS- and RBF-based polynomial predictors. Van Gestel *et al.* [31] investigated a Bayesian method to parameterize LS-SVMs for financial time series forecasting, especially in scenarios where the noise levels are comparable to the underlying signal energy and forecasting of the second moment (volatility) is needed. Cao *et al.* [32] developed a dynamic SVM model that uses an exponentially increasing regularization constant and an exponentially decreasing tube size to deal with structural changes in the data.

2.2.4 Hidden or State-observer Markov Models

Most of the models reviewed above involve batch processing, where the model is fitted and updated intermittently using batches of historical data. However, the curse of dimensionality due to the prohibitive computational effort, memory requirements, and large data sizes hampers their applicability to many real-world problems, especially for online process monitoring. A variety of sequential (also known as online or recursive) forecasting models, such as hidden Markov models (HMMs) [33], have been investigated to surmount this limitation. Some HMMs have been attempted for nonlinear time series forecasting, such as extended Kalman filters (EKF) [34] and particle filter (PF) models [35, 36].

2.3 Nonparametric Models

Although parametric models can provide accurate forecasts when the models are correctly specified, they tend to become highly suboptimal whenever the underlying dynamics are unknown or indeterminable; additionally, the issue of model biases persists as the dynamics of

most real-world complex systems are inherently nonlinear and nonstationary. In contrast, nonparametric models can provide a complete representation of the dynamics based on the observed data, do not impose any structural assumptions, and simplify modeling efforts. Consequently, the modeling and forecasting accuracy for nonlinear and nonstationary time series is improved. However, compared to parametric models, nonparametric models usually require larger datasets from which information on the underlying relationships may be effectively gleaned. Three classes of nonparametric models for nonlinear and nonstationary time series forecasting are reviewed in this section: state space neighborhood, Bayesian nonparametric, and functional decomposition models,

2.3.1 State Space Neighborhood and Local Topology-based Models

State space neighborhood approaches predict future values by selectively resampling historical observations, with the basic assumption that future behavior varies smoothly, i.e., observations similar to the target one are likely to have similar outcomes. Those models are attractive for complex system dynamics forecasting because of their simplicity and accuracy [37].

In Mehrotra *et al.*'s [38] k nearest neighbor (KNN) resampling approach with multiple predictor variables, an influence weight was assigned to each predictor to identify nearest neighbors. Hamid *et al.* [39] investigated a variety of neighbor-based methods for the forecasting of chaotic time series, e.g., zeroth-order approximation (one nearest neighbor), k -nearest neighbors (multiple neighbors) and weighted distance approximation (distance weighted average of multiple neighbors) models.

For most complex dynamic systems, it is not possible to observe all relevant variables. Often, the time evolution of only certain variables is observed, but the relationship to the state variables is unknown or indeterminable. The state space reconstructed from time delay embedding holds strong similarities (i.e., a diffeomorphic image) to the underlying state space [40] and offers a

unique way for nonlinear time series forecasting [41]. Casdagli [42, 43] developed a local linear model of the reconstructed state space for chaotic time series forecasting. The predicted value of the current observation was derived from the most recent w embedding vectors. Then k nearest neighbors were identified within the window width w based on the recurrence property of the reconstructed state space. Yang *et al.* [44] partitioned the reconstructed state space into various near-stationary segments based on local recurrence properties. The evolutionary trajectories were further decomposed using a principal component approach to identify the principal evolution directions and thus predict future states. Similarly, Bukkapatnam *et al.* [45] developed a local Markov model (LMM) based on the recurrence properties. A Markov transition matrix derived from the reconstructed state space was applied for system pattern analysis and to partition the trajectories into piecewise stationary segments. A local Markov model was employed to predict future states in each near-stationary segment based on the obtained transition matrix. McNames [46] studied time series forecasting with little or no noise. Neighbors in the reconstructed state space were identified by an optimized, and weighted Euclidean metric, and a novel ρ -step-ahead cross-validation error was used to assess model accuracy. Regonda *et al.* [47] investigated a local polynomial regression model using neighbors and future evolutions in the reconstructed state space. An ensemble nearest-neighbor model was then implemented via selecting a suite of parameter combinations for the local regression model. This ensemble approach was able to adequately capture the effects of parameter uncertainty.

2.3.2 Bayesian Nonparametric Models

The second class of nonparametric models is a group of Bayesian models. Bayesian modeling is basically a process of incorporating prior information to render posterior inference, i.e., estimating the conditional distribution $p(\theta|y)$ of the hidden model or parameters θ given an observed time series $y(t)$ [48]. Different from other Bayesian methods, Bayesian nonparametric models assume the hidden structure here grows with the data. In other words, Bayesian

nonparametric models seek a single model from infinitely many possibilities (i.e., θ may be infinite-dimensional) whose complexity (i.e., the number of parameters to estimate) is adapted according to the data. Among Bayesian nonparametric models, Gaussian process (GP) models have been the most widely studied for time series forecasting [49]. A GP model provides not just a point estimate, but a complete distribution of the forecasting. However, GP models have two major limitations, namely, the computational expense to perform the matrix inverse and the stationary covariance function assumption. Many attempts to address these issues have been investigated in the literature. Of these efforts, nonstationary covariance functions introduced to overcome the stationarity assumptions [50, 51] are suited only for simple nonlinear and nonstationary forms, such as linear trends, and require additional parameters to fit.

2.3.3 Functional Decomposition Models

Among nonparametric models for nonlinear and nonstationary forecasting, functional decomposition models have recently been receiving attention in the literature. The advantages of this type of models include local characteristic time scales and the use of an adaptive basis that does not require a parametric functional form. These models can be used to capture drifts and nonlinear modes of any nonlinear and nonstationary processes. Most of the models in this category are mixed or hybrid models that utilize a decomposition technique. Among nonparametric decomposition techniques, empirical mode decomposition (EMD) [52] can decompose nonstationary time series into a finite number of components called intrinsic mode functions (IMFs), such that evolutions of each IMF can be explored individually using different time scales via classical time series forecasting techniques, such as AR or ARMA models [53, 54]. Because EMD allows perfect reconstruction of original time series with IMFs and isolation of trend and noise components from a nonstationary process [55], it can improve long-term forecasting accuracy. Sa-ngasoongsong and Bukkapatnam [56] developed a two-step EMD model and applied it to long-term customer willingness-to-pay forecasting. An et al.'s [57] wind farm

power forecasting model was based on estimating the largest Lyapunov exponents for each IMF. They applied the largest Lyapunov exponent forecasting method [58] to provide higher forecasting accuracy than other direct EMD-based forecasting models. SVR models built on each IMF were also fused in different ways [59-61] to forecast nonstationary time series. EMD has also been applied for weather time series forecasting [62, 63]. Although attractive for nonlinear and nonstationary forecasting, EMD poses some mathematical challenges due to the edge effects [64]. Several possible fixes such as adaptive decompositions have been attempted in the context of financial time series forecasting [65, 66].

Recently, intrinsic time-scale decomposition (ITD) has been investigated for precise time-frequency-energy (TFE) analysis of time series [2]. ITD overcomes the limitations of classical Fourier, wavelet, and EMD approaches for nonlinear and nonstationary time series modeling and decomposes time series into proper rotation components with defined frequency and amplitude and a monotonic trend. This decomposition preserves precise temporal information regarding critical points and riding waves in time series, with a temporal resolution equal to the time-scale of extrema occurrence of the input signal. The ITD model has been used for nonlinear biomedical signal characterization [67], and forecasting long-term automobile demand has been attempted using the ITD-based vector error correction model (VECM) [68].

CHAPTER III

RESEARCH METHODOLOGY

This chapter presents a research methodology for developing an effective prediction model to forecast real-world nonlinear and nonstationary time series. In this study, two advanced nonlinear decomposition techniques, empirical mode decomposition (EMD) and intrinsic time-scale decomposition (ITD), were used to develop new approaches to forecasting nonlinear and nonstationary time series. EMD-based prediction methodology was specifically developed for univariate time series prediction. This method addresses the challenge (edge artifacts) of using EMD for prediction applications. ITD-based prediction methodology takes the advantage of structural relationship identification (SRI) methodology to identify endogenous and exogenous variables for multivariate time series prediction. The overall research methodology in this chapter consists of three main parts: EMD-based prediction methodology, SRI methodology, and ITD-based prediction methodology.

3.1 Empirical Mode Decomposition (EMD) based Prediction Methodology

This section presents a new prediction approach based on the EMD technique. EMD is an advanced nonlinear decomposition technique that aims to decompose a nonstationary signal into a finite number of components called intrinsic mode functions (IMFs). EMD allows perfect reconstruction of the original signal using IMFs. This property of EMD leads to an application of determining trend and noise from time series. However the presence of edge artifacts limits

the use of EMD for prediction applications. A key aspect to this approach is that it takes advantage of the linear phase property of the Hilbert-Huang transform (HHT) to address the edge artifact limits, thus extending EMD for long-term prediction applications.

3.1.1 Empirical Mode Decomposition

Empirical mode decomposition (EMD), first proposed by Huang [3], represents a nonlinear, nonstationary signal $x(t)$ as the superposition of a finite number of components called intrinsic mode functions (IMFs) as

$$x(t) = \sum_{n=1}^L s_n(t) + r_{L+1}(t) \quad (3.1)$$

where $s_n(t)$ are the IMFs and $r_{L+1}(t)$ is the residual from the decomposition. The IMFs provide local characteristic time scales and an adaptive basis for EMD. Consequently, IMFs can be used to capture drifts and nonlinear modes of any nonlinear and nonstationary processes. Also, the adaptive basis of EMD does not require a parametric functional form for time series. The characteristic of IMF is that it has only one extreme between zero crossings and has a mean value of zero. The following EMD procedure, called the sifting process, is used to extract IMFs from a time series $x(t)$:

(1) Identify all local minima and maxima ($x_{\max}^*(t_i)$ and $x_{\min}^*(t_j)$) of $x(t)$, where $i = 1, \dots, n$ and

$$j = 1, \dots, m$$

(2) Use a cubic spline interpolation to define an upper $u(t)$ and lower envelope $v(t)$ from the extreme points [69] as

$$u(t) = \oplus u_i(t) \quad t \in (t_i, t_{i+1}) \text{ and } i = 1, \dots, n \quad (3.2)$$

where \oplus represents a direct sum,

$$u_i(t) = \frac{z_{i+1}}{6\tau_i}(t-t_i)^3 + \frac{z_i}{6\tau_i}(t_{i+1}-t)^3 + C_i(t-t_i) + D_i(t_{i+1}-t) \quad (3.3)$$

$$\tau_i = t_{i+1} - t_i, C_i = \frac{x_{\max}^*(t_{i+1})}{\tau_i} - \frac{\tau_i}{6} z_{i+1}, D_i = \frac{x_{\max}^*(t_i)}{\tau_i} - \frac{\tau_i}{6} z_i \quad (3.4)$$

$$\tau_{i-1} z_{i-1} + 2(\tau_{i-1} + \tau_i) z_i + \tau_i z_{i+1} = 6 \left(\frac{x_{\max}^*(t_{i+1}) - x_{\max}^*(t_i)}{\tau_i} - \frac{x_{\max}^*(t_i) - x_{\max}^*(t_{i-1})}{\tau_{i-1}} \right) \quad (3.5)$$

The process for defining the lower envelope $v(t)$ is similar to Eqs. (3.2)-(3.5) when

$$v(t) = \oplus v_j(t) \quad t \in (t_j, t_{j+1}) \text{ and } j = 1, \dots, m$$

(3) Compute the mean envelope $m(t)$

$$m(t) = [u(t) + v(t)] / 2 \quad (3.6)$$

(4) Compute the IMFs $s_p(t)$ iteratively through a series of reductions

$$h_k(t) = r_{p-1}(t) - m_k(t), \quad k = 1, \dots, K_{\max}, p = 1, \dots, P, \quad r_0(t) = x(t) \quad (3.7)$$

Here, $h_k(t)$ is treated as the series, and $m_k(t)$ is computed as the mean of the upper and lower envelopes of $h_k(t)$. The process repeats all the steps until the following stoppage criterion (SD) is reached

$$\frac{\sum_{t=0}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^T h_{k-1}^2(t)} < SD^* \quad (3.8)$$

(5) Obtain the residue $r_p(t)$

$$r_p(t) = r_{p-1}(t) - s_p(t) \quad (3.9)$$

The process continues till the number of the extreme points of $r_p(t)$ is not larger than two. The Hilbert-Huang transform (HHT) may be used to represent the time-frequency content of the IMFs as [70]

$$s_n^a(t) = A_n(t) e^{j\phi_n(t)} \quad (3.10)$$

Where amplitude envelope $A_n(t) = |s_n^a(t)| = \sqrt{s_n^2(t) + \bar{s}_n^2(t)}$ and instantaneous phase

$\phi_n(t) = \arg\{s_n^a(t)\}$. The following property of the instantaneous phase provides some key benefits

for long-term prediction.

Remark 1: The instantaneous phase $\phi_n(t)$ extends linearly over time, i.e., $\phi_n(t) \approx \phi_0 + \alpha t$.

Consequently, if the amplitude envelope is predicted, one can develop an extended prediction of the series. Fig. 3.1 shows an example of customer willingness-to-pay (WTP) for the fuel economy attribute of an automobile.

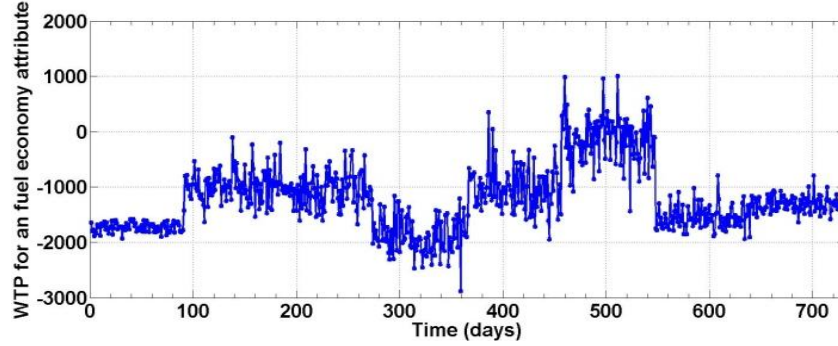


Figure 3.1: Customer WTP for a fuel economy attribute

Fig. 3.2 shows the amplitude envelope and instantaneous phase of IMF1 of WTP (shown in Fig. 3.1) using the EMD technique. In the case of higher order IMFs (See Fig. 3.3), amplitude envelopes tend to change slowly over time. Fig. 3.3 shows the amplitude envelop of IMFs 2-9 for WTP for the fuel economy attribute.

Although, EMD provides a convenient representation of data from nonlinear and nonstationary processes, the following issue arising from the mathematical construct of the EMD algorithm hinders its application to nonstationary process prediction. The cubic splines used in the EMD process are very sensitive to the end points (edge effects) over which they are parameterized. The edge effects of EMD can significantly affect the estimation of the mean function, $m(t)$, and the IMFs. Several possible fixes [70] have been studied to address current algorithms for the edge effects issue. They mostly involve extending the record of maxima and minima. However, most of the algorithms reported in the literature, such as statistical extrapolation and mirror image extension [52, 71], do not consider the true extension of the signals. Consequently, addressing the

edge effects issue remains an area of active research. To address this limitation, a two-step approach for EMD-based prediction is presented.

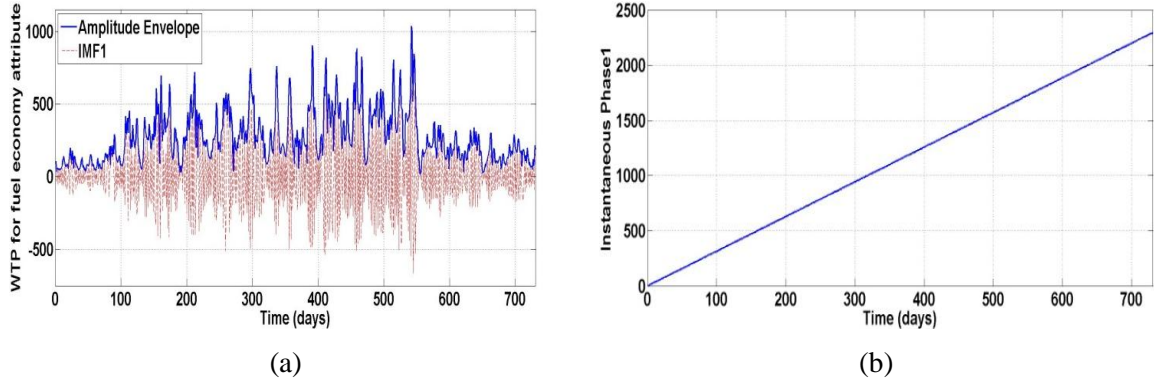


Figure 3.2: (a) Amplitude Envelope and (b) Instantaneous Phase of IMF1

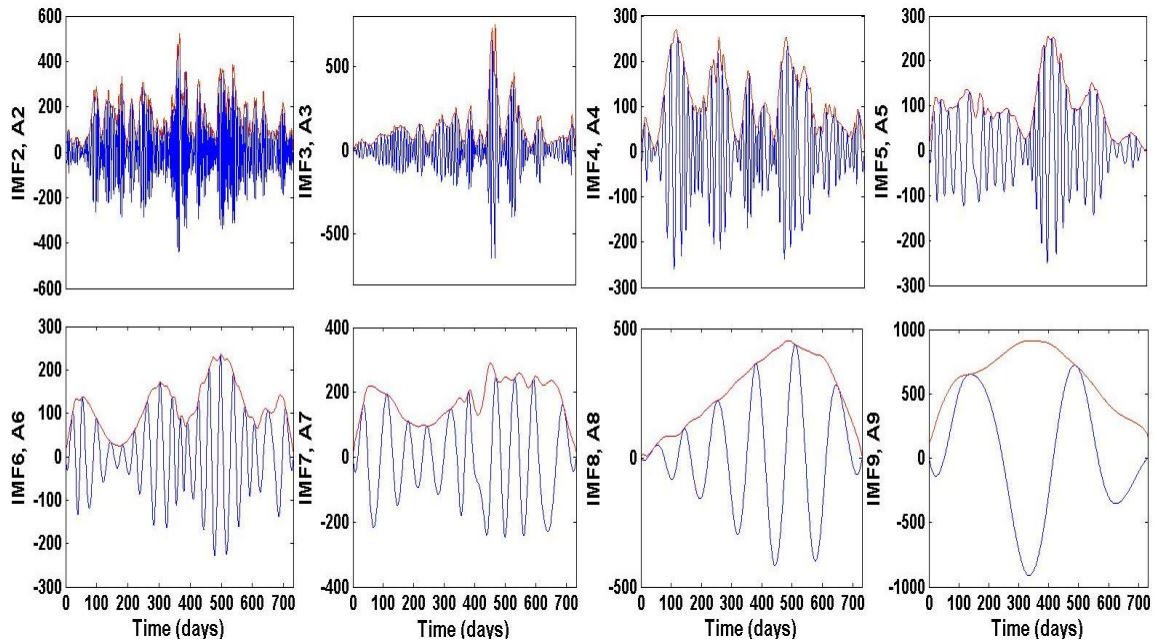


Figure 3.3: Amplitude Envelope of IMFs 2-9

3.1.2 Two-step Approach for EMD-based Prediction

The prediction approach based on EMD consists of two steps. The first step involves estimating the long-term trend component and reducing noise components in the time series signal [72]. In the second step, the prediction results from the first step are then used to reduce the edge effects

of EMD by decomposing the concatenation of the original time series and the first step predictions to obtain new IMFs and the residue. The prediction based on linear extension of the instantaneous phase is then applied to each of the new IMFs and the residue. The final prediction values are obtained using reconstruction of the signal from the predictions of all new IMFs and the residue.

3.1.2.1 The First Step

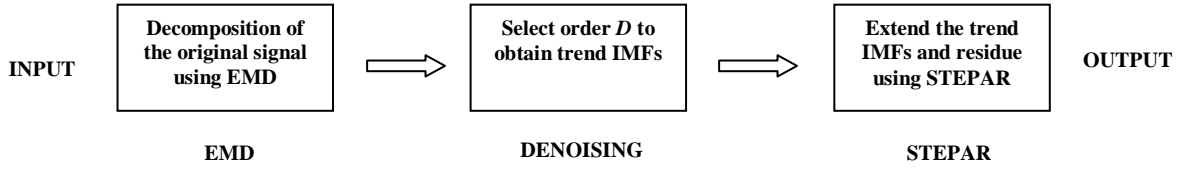


Figure 3.4: The First Step Procedure

The procedure for the first step of an algorithm is shown in Fig 3.4. EMD allows a perfect reconstruction of the original series, $x(t)$, using IMFs and the residue as given in Eq. (3.1). This property of EMD allows the estimation of both trend and noise (detrending and denoising) in a nonstationary process. Wu and Huang [73, 74] proposed a method to assign a statistical significance to the information content for IMF components derived from noisy data using the relationship between the energy density and the average period of IMFs. They defined a spread function using white noise references of identical length with a target dataset. If the energy density of IMFs lies outside the upper and lower bounds of the spread functions, those IMFs will be treated as containing information. Flandrin *et al.* [72] used an elegant method to remove undesirable components by partially reconstructing the original series, $x(t)$, as shown in the following function:

$$\hat{x}_D(t) = \sum_{n=1}^D s_n(t) \quad (3.11)$$

where D is the largest IMF index considered to be noise IMF. In this approach, the IMFs 1 to D are considered to be noise. Higher index IMFs are considered as containing information.

Theoretically, IMFs should have a mean value of zero. However, the average of an IMF may have a value that may be a small non-zero quantity. To select D , each of the IMFs is initially estimated. Then the series of average IMFs is standardized with $\mu=0$ and $\sigma=1$. A rule for choosing D is to observe the evolution of the (standardized) empirical mean of $\hat{x}_d(t)$ as a function of a test order d , and to identify for which $d = D$ it departs significantly from zero. Fig. 3.5 illustrates the procedure to select order d of the IMF (WTP data) from Eq. (3.11). The order $d = 7$ (IMF 7) is selected as the change point. Therefore, the detrended series in Fig. 3.5(c) is obtained from the partial reconstruction with IMFs 1 to 7.

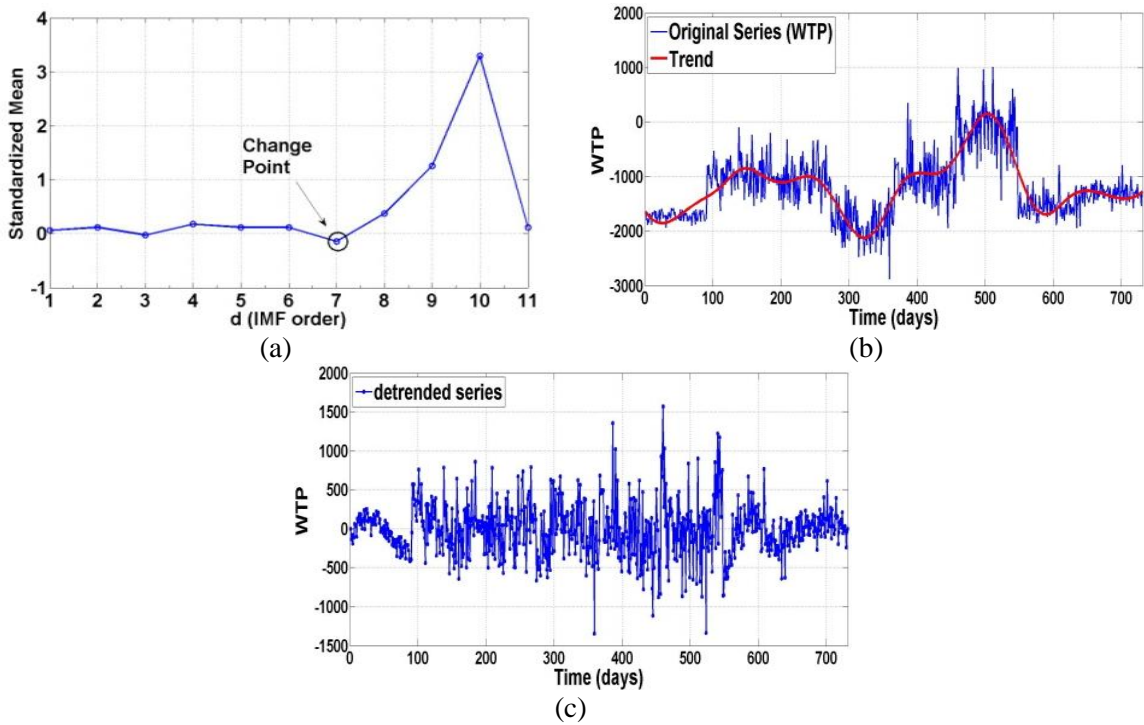


Figure 3.5: Detrending operation by observing the change point. (a) Standardized mean for each IMF order. (b) Trend and original series (WTP for the fuel economy attribute). (c) Detrended series from detrending operation.

As shown in Fig. 3.5, the trend component of WTP for the specified attribute consists of IMFs 8 through 11 and the residue. The original IMFs and residue are individually extended using the stepwise autoregressive method. This STEPAR initially fits an autoregressive model and then

removes all autoregressive parameters with large p-values. The STEPAR(q) model can be represented as

$$s_n(t) = a_0 + \sum_{i=1}^{q^*} a_i s_n(t-i) + \varepsilon_t \quad (3.12)$$

where lags order $i=1$ to q^* are sequentially selected using a stepwise procedure. In this two-step approach, the maximum number of autoregressive parameters (lag order) is set to twelve.

3.1.2.2 The Second Step

In the second step, the predicted trend component is reconstructed from the predicted IMF components using Eq. (3.11). The predicted trend is concatenated with the original data to obtain the extended signal. New IMFs and residue are obtained using EMD on the extended signal. The number of new IMF components can differ from that of the original signal, depending upon the length of the prediction step, and new local minima and maxima in the predicted trend. Irrespective of the edge effects issue, the prediction of the trend component in the first step can be considered a variant of classical time series techniques where a separate analysis is applied to each IMF component. The result of the first step can significantly outperform classical and advanced time series techniques for a nonstationary series whose long-term trend is nonlinear and which exhibits multiple frequency content. However, the edge effects problem of EMD limits the use of IMFs for prediction as shown in Fig. 3.6(a). All IMFs begin and end with zero value. With the second step procedure on WTP data, 13 new IMFs and the residue can be obtained without these edge artifacts. Fig. 3.6(b) shows the new trend IMFs from the first step procedure.

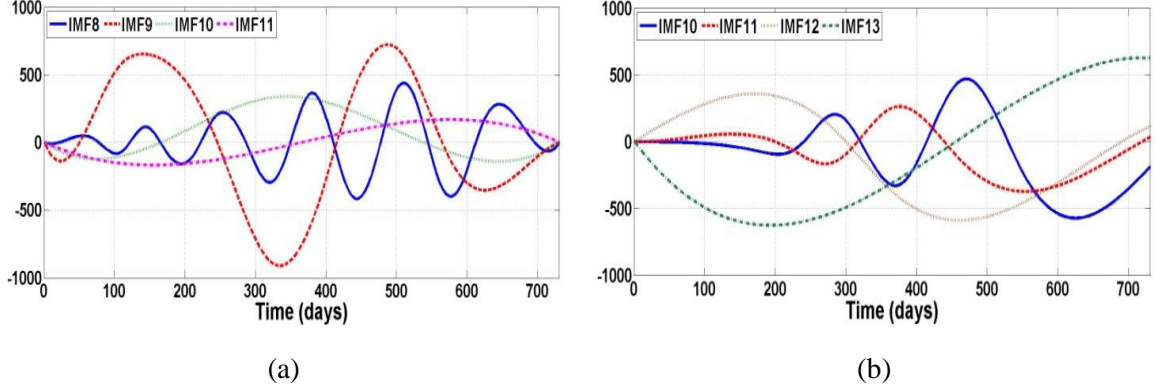


Figure 3.6: (a) Original trend IMFs and (b) New IMFs from the first step procedure

A noteworthy issue in using IMFs for prediction is that some IMFs may be more influential than others for long-term prediction. These are usually the IMFs with the longest range and not necessarily the highest order IMFs (e.g., IMF 9 in Fig. 3.6(a) and IMF13 in Fig. 3.6(b)). To avoid the effect of these dominant IMFs, the prediction horizon should be restricted to the period where prediction results are equally influenced by each trended IMF. Based on the definition of EMD, Wu and Huang [73] defined the average period (\bar{T}_n) of each IMF calculated from any given spectrum of the Fourier spectrum function for the n^{th} IMF. This average period value is almost identical to the number of points divided by the number of peaks (local maxima) or zero crossings. Since the focus is on long-term prediction, the suggested prediction horizon in this study is one fourth of the averaged period of the lowest-order trend IMF as stated in the following:

Remark 2: The effective prediction horizon for p^{th} IMF may be estimated as

$$\hat{Q}_p^{\max} = \frac{1}{4} \int S_{\ln T, p} d\ln T \left(\int S_{\ln T, p} d\ln T \frac{d\ln T}{T} \right)^{-1} \quad (3.13)$$

where $S_{T, p}$ is the Fourier spectral component of period T for the p^{th} IMF.

For example, IMF 8 is the lowest-order trend IMF with the averaged period of 56.23 (number of zero crossings = 13). Therefore the suggested prediction horizon is 14.06 (≈ 14). As mentioned in

Section 3.1, the advantage of the instantaneous phase is that it extends linearly over time which facilitates long-term predictability. However, amplitude envelope prediction must also be accounted for to receive an accurate prediction of the series. In the second step of the prediction algorithm, amplitude envelope prediction is received from computing the amplitude of new IMFs and the residue from the first step results. The linear extension of the instantaneous phase is then applied to each of the new IMFs and residue. The prediction equation is shown in Eq. (3.14).

$$\hat{x}(t) = \text{Re} \sum_{n=1}^{L+1} \hat{A}_n(t-p) e^{j\hat{\phi}_n(t)} \quad (3.14)$$

where Re is the real part, $p = 1$, $\hat{\phi}_n(t) = \phi_0 + \alpha t$, and $L+1$ order represents the residue term of EMD.

3.2 Structural Relationship Identification (SRI) Methodology

As one of the research objectives is to develop an effective prediction model to forecast multivariate, nonlinear, and nonstationary time series, identification of the interrelationships among time series variables in the system is absolutely essential. In the econometric time series research area, some developments in multivariate time series techniques have been specifically designed to quantify the structural relationship of variables in the system. These models are vector autoregressive (VAR) and vector error correction models (VECM) [75, 76]. They have been broadly recognized as powerful, theory-driven models that can be used to describe the long-run dynamic behavior of multivariate time series. Especially in the case of nonstationary variables with cointegration, the VECM provides the best persistent modeling for addressing the problem of long-run market-response identification by estimating the cointegration vectors of the endogenous variables. However, there is some difficulty associated with this type of model. The identification of the structural relationship of the variables of these models is often problematic. For example, the large number of endogenous and exogenous variables may cause an over-

parameterization issue leading to wrong estimations of all model coefficients. Model misspecification in VAR and VECM may also occur due to introducing irrelevant endogenous variables into the models.

This section presents a structural relationship identification (SRI) methodology to address these pertinent issues in VAR and VECM. The methodology presented in this section includes a battery of advanced statistical methods to test for a unit root, weak exogeneity, Granger causality, and cointegration. The unit root test identifies the stationarity of each variable in the model. A sequential reduction method of weak exogeneity addresses an over-parameterization problem and also identifies the reasonable structural relationships of endogenous variables. Granger causality identifies causal relationships for the variable selection process. Because of the nonstationary characteristic of the variables, a cointegration test is used to identify the existence of long-run equilibrium relationships among selected endogenous variables.

Fig. 3.7 presents an example of automobile sales modeling using the SRI methodology framework. In this section, the procedures used to identify the structural relationship of variables in the system are described as follows:

The first step is to apply unit root tests [77-80] to identify the stationarity of each variable in the dataset. This step is also used to identify the order of integration for nonstationary variables. VAR or VARX (vector autoregressive with exogenous variables) models [75, 81, 82] can be applied if all variables in the dataset are stationary.

Weak exogeneity [83-85] and Granger-causality tests [86-88] are used to determine the exogeneity of each variable in the second step. A sequential reduction method for weak exogeneity [89] is to apply a sequence of weak exogeneity tests by exogenizing any non-rejecting weakly exogenous variables and re-testing the remaining variables until all weakly exogenous variables are identified. For nonstationary variables, the cointegration rank test is used to

determine whether a long-run relationship among endogenous variables exists. Cointegration typically refers to a linear combination of nonstationary endogenous variables in the system [79, 81, 83, 87, 90, 91]. VECM or VECMX (vector error correction model with exogenous variables) models can be applied if cointegration exists among endogenous variables, and VAR or VARX models in differences can be applied if there is no cointegration in the system.

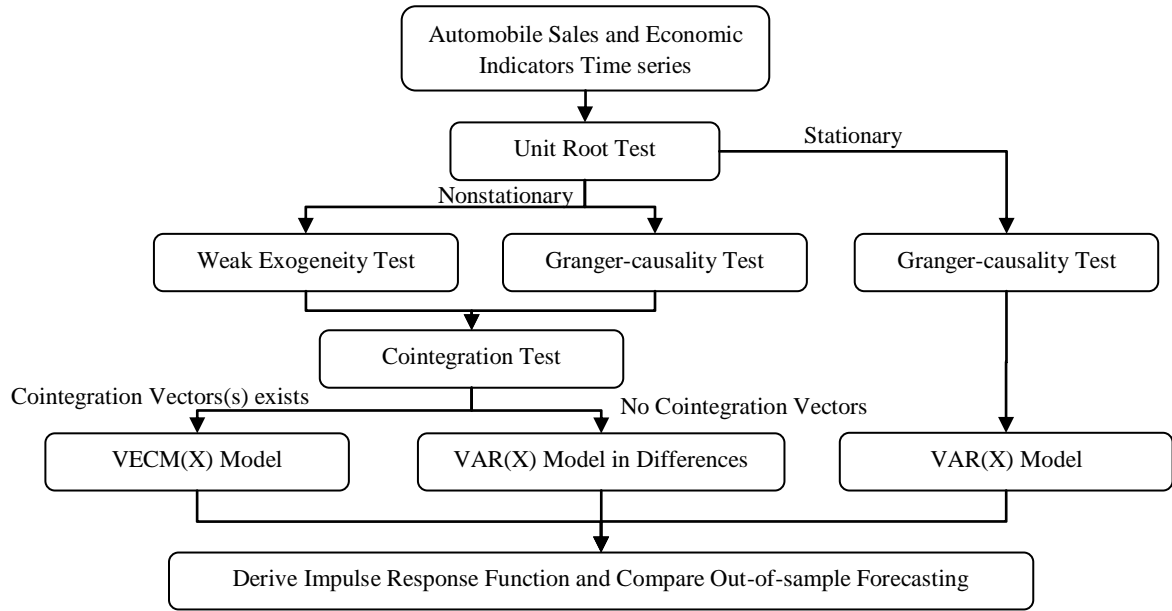


Figure 3.7: Automobile Sales Modeling Framework (SRI Methodology)

The final step in the framework is to compare the out-of-sample forecasting performance and derive the impulse response function from the selected model to trace the effect of a one-unit change in one of the variables on the future values of endogenous variables. A brief overview of each test and model used in the present approach is presented in Sections 3.2.1 through 3.2.5. Since many of these tests and modeling methods are well established in the literature (as summarized in the foregoing), only the necessary overview is included so that a consistent notation and symbology are established.

3.2.1 VAR and VECM

Vector autoregression (VAR) is a natural extension of the univariate autoregression (AR) models. It treats each variable in the system symmetrically. The evolution of each endogenous variable in the system can be explained by its own lags and the lags of all other endogenous and exogenous variables (in the case of VARX). VAR can also handle feedback from each endogenous variable, feedback which can be used to explain the structural relationships of variables in the system. Let $y_t = (y_{1t}, \dots, y_{nt})'$, $t = 1, 2, \dots$ denote an n -dimensional time series vector. A p^{th} -order vector autoregression, denoted as VAR(p), can be represented as

$$\begin{aligned} \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} &= \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} + \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} & \dots & A_{1n}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} & \dots & A_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}^{(1)} & A_{n2}^{(1)} & \dots & A_{nn}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{nt-1} \end{bmatrix} + \dots + \begin{bmatrix} A_{11}^{(p)} & A_{12}^{(p)} & \dots & A_{1n}^{(p)} \\ A_{21}^{(p)} & A_{22}^{(p)} & \dots & A_{2n}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}^{(p)} & A_{n2}^{(p)} & \dots & A_{nn}^{(p)} \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \\ \vdots \\ y_{nt-p} \end{bmatrix} \\ &+ \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \end{aligned} \quad (3.15)$$

where c_i = parameters representing intercept terms, $i = 1, 2, \dots, n$.

$A_{ij}^{(s)}$ = autoregressive coefficients, $i = j = 1, 2, \dots, n$ and $s = 1, 2, \dots, p$.

ε_{it} = white noise disturbances, $i = 1, 2, \dots, n$

A VAR process can also be affected by exogenous variables. In vector autoregression with exogenous variables (VARX), the VARX(p, m) can be represented in a simple form as

$$y_t = c + \sum_{l=1}^p \Phi_l y_{t-l} + \sum_{l=0}^m \Theta_l^* x_{t-l} + \varepsilon_t \quad (3.16)$$

where c denotes an $(n \times 1)$ vector of constants and Φ_l denotes an $(n \times n)$ matrix of autoregressive coefficients for $l = 1, 2, \dots, p$. The $(n \times 1)$ vector ε_t is a vector with an Ω $(n \times n)$ symmetric positive definite matrix. The $x_t = (x_{1t}, \dots, x_{wt})'$ is a w -dimensional exogenous time

series vector and Θ_l^* is an $n \times w$ matrix of coefficients. For stationary assumptions of VAR models, the stationary condition is satisfied if all roots of $|\Phi(z)| = 0$ lie outside the unit circle. If the stationary condition is not satisfied, a nonstationary model (a differenced model or an error correction model) might be more appropriate.

The vector error correction model (VECM) is a dynamic multi-factor system that adds error correction features into the VAR model. A special characteristic of this model is that we can identify a long-run equilibrium relationship which can be used to improve longer-term forecasting of variables in the system. This long-run equilibrium relationship can be determined from the cointegration vector. The vector error correction model with the cointegration rank $r (\leq n)$, denoted as VECM(p), can be written as

$$\Delta y_t = \delta + \Pi y_{t-1} + \sum_{l=1}^{p-1} \Phi_l^* \Delta y_{t-l} + \varepsilon_t \quad (3.17)$$

where r is the number of cointegrating vectors; Δ is the differencing operator, such that $\Delta y_t = y_t - y_{t-1}$; $\Pi = \alpha\beta'$, where α and β are $n \times r$ matrices; and Φ_l^* is an $n \times n$ matrix. The cointegrating vector, β , is sometimes called a long-run parameter, and α is an adjustment coefficient. In the case of cointegration with exogenous variables, the VECM with exogenous variables, VECMX (p, m), can be written as follows:

$$\Delta y_t = \delta + \Pi y_{t-1} + \sum_{l=1}^{p-1} \Phi_l^* \Delta y_{t-l} + \sum_{l=0}^m \Theta_l^* x_{t-l} + \varepsilon_t \quad (3.18)$$

Theoretically, VECM should be considered if and only if y_t is cointegrated, as defined in the following definition:

Definition 1: The components of the vector $y_t = (y_{1t}, \dots, y_{nt})'$ are said to be cointegrated of order d, b , denoted by $y_t \sim CI(d, b)$ if (I) components of y_t are integrated of order $d, I(d)$, and

(II) there exists a vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ such that the linear combination $\beta y_t = \beta_1 y_{1t} + \beta_2 y_{2t} + \dots + \beta_n y_{nt}$ is integrated of order $(d - b)$, $I(d - b)$, where $b > 0$.

To identify the structural relationship of variables in the system, including causality and cointegration, the following subsections provide the details of each statistical hypothesis test involving SRI methodology.

3.2.2 Unit Root Tests

Two unit root tests are used to test the stationarity of the variables in the system,. The procedure implemented here is to test the null hypothesis of a unit root against the alternative of stationarity. The first test is the augmented Dickey and Fuller (ADF) [78, 92] unit root test. The ADF test provides the unit root test for three specific cases: a unit root with zero mean, unit root with drift, and unit root with drift and a deterministic linear time trend. For the unit root test with nonlinear trend stationarity, the second test, the univariate G-test [80] is applied. According to Park [79], insufficient attention has been given to the specification of the deterministic trend in the unit root test. The univariate G-test is distinct from conventional unit root tests in that it includes the nonlinear polynomial trend in the equations, resulting in a unit root test with nonlinear trend stationarity.

This version of the ADF unit root test is the augmented version of the Dickey-Fuller test. Dickey and Fuller [78, 92] showed that the test statistic τ of the Dickey-Fuller test does not follow a t -distribution. The two versions of the test used in this paper are shown in Eq. (3.19) and (3.20).

$$\Delta y_t = \alpha + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (3.19)$$

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (3.20)$$

The null hypothesis is $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$. Eq. (3.19) is the test for a unit root with drift and Eq. (3.20) is a test for a unit root with drift and a deterministic time

trend. The appropriate lag length for the ADF test is selected using the general-to-specific methodology [93]. The idea is to select the optimal lag (p^*) that is significantly different from zero for the p^{th} autoregressive process shown in Eqs. (3.19) and (3.20).

For the unit root test with nonlinear trend stationarity, Park and Choi [79] proposed the univariate G-test specifying ordinary time polynomials in the deterministic time trend. The univariate G-test is a variable addition test, since it is based on the regression of a given time series on time polynomials. To test whether a given time series is $I(0)$ ¹ or stationary, the series is regressed on a time polynomial with order dictated by the null hypothesis and then some higher-order time polynomial terms are added. Suppose a time series y_t can be decomposed as

$$y_t = \sum_{\tau=0}^p \mu_{\tau} t^{\tau} + \sum_{\tau=p+1}^q \mu_{\tau} t^{\tau} + \eta_t \quad (3.21)$$

Let $0 \leq p < q$ and consider the following least squares regressions

$$y_t = \sum_{\tau=0}^p \mu_{\tau} t^{\tau} + \tilde{e}_{(p)t} \quad (3.22)$$

$$y_t = \sum_{\tau=0}^q \mu_{\tau} t^{\tau} + \tilde{e}_{(q)t} \quad (3.23)$$

and define $\hat{\sigma}_{(p)}^2 = \frac{1}{T} \sum_{t=1}^T \tilde{e}_{(p)t}^2$ and $\hat{\sigma}_{(q)}^2 = \frac{1}{T} \sum_{t=1}^T \tilde{e}_{(q)t}^2$

The test statistic is constructed by normalizing the standard Wald statistic as

$$G(p, q) = \frac{T(\hat{\sigma}_{(p)}^2 - \hat{\sigma}_{(q)}^2)}{\hat{\omega}^2} = \frac{\hat{\sigma}_{(q)}^2}{\hat{\omega}^2} F(\alpha_{p,q}) \quad (3.24)$$

¹The $I(d)$ process requires d^{th} difference to be stationary. A stationary process is denoted as an $I(0)$ process.

where $F(\alpha_{p,q}) = \frac{T(\hat{\sigma}_{(p)}^2 - \hat{\sigma}_{(q)}^2)}{\hat{\sigma}_{(q)}^2}$ is the standard Wald statistic for the null hypothesis of stationarity

(I(0)) that $\tilde{\mu}_{p+1} = \tilde{\mu}_{p+2} = \dots = \tilde{\mu}_q = 0$, and $\hat{\omega}^2$ is the long-run variance of η_t .

$$\hat{\omega}^2 = \frac{1}{T-1} \sum_{t=2}^T \hat{\eta}_t^2 + 2 \sum_{j=1}^l \left(1 - \frac{j}{l+1}\right) \frac{1}{T-1} \sum_{t=2}^T \hat{\eta}_t^2 \hat{\eta}_{t-j}^2 \quad (3.25)$$

It is shown by Park and Choi (1988) that $G(p, q) \rightarrow \chi_{q-p}^2$ if y_t is I(0) process, while $G(p, q) \rightarrow \infty$ if y_t is an I(1) process.

3.2.3 Weak Exogeneity Test

The weak exogeneity of variables indicates that no useful information is lost when these variables are conditioned without specifying their generating process. For example, in the case of the automobile sales modeling, the two automobile sales (Y_i 's) are hypothesized as endogenous variables and the remaining variables (economic indicators, X_i 's) as exogenous variables. The test of weak exogeneity will identify the weak exogeneity effect of each variable on the others. Testing the weak exogeneity assumption also addresses the over-parameterization problem of VAR and VECM modeling, so that the number of equations in the model can be reduced if the variables are treated as weakly exogenous. To test which variables should be treated as endogenous in the equation and which ones as exogenous, the n -vector of I(1) random variables y_t is initially partitioned into the n_1 -vector y_{1t} and the n_2 -vector y_{2t} , where $y_t = (y'_{1t}, y'_{2t})$ and $n = n_1 + n_2$. From the VECM model (see Section 3.1), the parameters can similarly be decomposed as $\delta = (\delta'_1, \delta'_2)$, $\alpha = (\alpha'_1, \alpha'_2)$, and $\Phi_i^* = (\Phi_{1i}^*, \Phi_{2i}^*)$ and the variance-covariance matrix as $\Sigma = [\Sigma_{11} \ \Sigma_{12}; \Sigma_{21} \ \Sigma_{22}]$. The conditional model for y_{1t} given y_{2t} is

$$\begin{aligned} \Delta y_{1t} = & \omega \Delta y_{2t} + (\alpha_1 - \omega \alpha_2) \beta' y_{t-1} + \sum_{i=1}^{p-1} (\Phi_{1i}^* - \omega \Phi_{2i}^*) \Delta y_{t-i} + (\delta_1 - \omega \delta_2) \\ & + (\varepsilon_{1t} - \omega \varepsilon_{2t}) \end{aligned} \quad (3.26)$$

and the marginal model for y_{2t} is

$$\Delta y_{2t} = \alpha_2 \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_{2i}^* \Delta y_{t-i} + \delta_2 + \varepsilon_{2t} \quad (3.27)$$

where $\omega = \Sigma_{12} \Sigma_{22}'$

The test of weak exogeneity of y_{2t} determines whether $\alpha_2 = 0$. If $\alpha_2 = 0$, y_{2t} is said to be weakly exogenous, i.e., y_{2t} can be estimated without the need for an error correction model because wherever $\alpha_2 = 0$ holds, it causes β to drop out of the marginal model for y_{2t} . In this study, the sequential reduction method of weak exogeneity suggested by Greenslade *et al.* [89] is presented. The idea is to apply a sequence of weak exogeneity tests by exogenizing any non-rejecting weakly exogenous variables and re-testing the remaining variables until all weakly exogenous variables are identified.

3.2.4 Cointegration Test

Engle and Granger system [79, 81, 83, 87, 90, 91] showed that if a linear combination of nonstationary time series is stationary, the time series are cointegrated. Cointegrated processes appear independent in the short term but tend to move together in the long term. For a test of cointegration, Johansen's reduced rank methodology [91, 94] is employed. The advantage of the Johansen test is that it allows a system to have more than one cointegrating relationship. This test is generally more applicable than the Engle and Granger test. Considering the VECM in Eq. (3.17), if the coefficient matrix Π has reduced rank $r < n$, then there exists $n \times r$ matrices α and β such that $\Pi = \alpha \beta'$ and $\beta' y_t$ is stationary. Johansen has shown that for a given r , the maximum likelihood estimator of β can define the combination of y_{t-1} that yields the r largest canonical correlations of Δy_t with y_{t-1} after correcting for lagged differences and deterministic variables when present.

Two Johansen test statistics for cointegration are suggested for testing the null hypothesis that there are at most r cointegrating vectors ($H_0: \lambda_i = 0$ for $i = r + 1, \dots, k$). The trace and maximum eigenvalue statistics are as follows:

$$\lambda_{trace} = -T \sum_{i=r+1}^k \log(1 - \lambda_i) \quad (3.28)$$

$$\lambda_{max} = -T \log(1 - \lambda_{r+1}) \quad (3.29)$$

where T is the sample size and λ_i is the i^{th} largest canonical correlation. The trace test tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of n cointegrating vectors. For the maximum eigenvalue test, it tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of $r + 1$ cointegrating vectors. Asymptotic critical values of these tests are given in [95].

3.2.5 Granger-causality Test

The concept of Granger causality is that there exists a subset of variables that are useful for forecasting others. By definition, y_t is said to Granger-cause x_t if y_t can be used to help predict some stage in the future of x_t . In the case of no cointegration among variables in the system, this concept can be tested for a selection of endogenous and exogenous variables in the VAR model. Considering the example of the bivariate VAR(p) model with coefficients $A_{ij}^{(s)}$ for $i, j = 1, 2$ and $s = 1, \dots, p$ as follows:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \dots + \begin{bmatrix} A_{11}^{(p)} & A_{12}^{(p)} \\ A_{21}^{(p)} & A_{22}^{(p)} \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (3.30)$$

The variable y_{1t} is said to cause (Granger) y_{2t} , but y_{2t} does not cause (Granger) y_{1t} if $A_{12}^{(s)} = 0 \ \forall s$. This model structure implies that if $A_{12}^{(s)} = 0$, y_{1t} is influenced only by its own past values and not by the past of y_{2t} . For a larger VAR model ($n > 2$), Granger causality can be used to test whether one variable is influenced only by itself and not by other variables in the model system. Let y_t be an n -dimensional time series vector which is arranged and partitioned in

subgroups y_{1t} and y_{2t} with dimensions n_1 and n_2 , respectively ($n = n_1 + n_2$): that is, $y_t = (y'_{1t}, y'_{2t})'$ with the corresponding white noise process $\varepsilon_t = (\varepsilon'_{1t}, \varepsilon'_{2t})'$. Equivalent representation of the VAR model in Eq. (5.1) can be written as

$$\begin{bmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (3.31)$$

Consider testing $H_0: CB = c$, where C is an $s \times (n^2p + n)$ matrix of rank s and c is an s -dimensional vector where $s = n_1 n_2 p$. Assume that $\sqrt{T}(\hat{B} - B) \xrightarrow{d} N(0, \Gamma_p^{-1} \otimes \Sigma)$, as in least square (LS) and/or maximum likelihood (ML) estimation. The Wald statistic can also be obtained from [96] as

$$\sqrt{T}(C\hat{B} - c)[C(\hat{\Gamma}_p^{-1} \otimes \hat{\Sigma})C']^{-1}(C\hat{B} - c) \xrightarrow{d} \chi^2(s) \quad (3.32)$$

If the null hypothesis cannot be rejected, we can exclude that variable from the VAR model or treat it as an exogenous variable in the VARX model. As shown in the modeling framework, the Granger-causality test can be applied to both stationary and nonstationary systems. However, for nonstationary systems with the same order of integration, the cointegration relationship must also be identified to avoid the problem of using the VAR for a cointegrating system.

3.3 Intrinsic Time-Scale Decomposition (ITD) based Prediction Methodology

This section presents a new prediction approach based on the intrinsic time-scale decomposition (ITD) technique. The ITD method is a recently developed nonparametric decomposition technique for signals that are nonlinear and/or nonstationary in nature. ITD decomposes a signal into a sum of components called proper rotation components and a monotonic trend. As one of the advantages of a nonparametric method, ITD requires very limited assumptions about the data. This advantage makes ITD suitable for nonlinear data from unknown underlying processes. ITD also provides unbiased decomposing components, compared to a parametric algorithm, because of parameter estimation. The key aspect of this new prediction approach is to use the SRI

methodology presented in Section 3.2 to identify a causal and long-run equilibrium relationship between variables and the ITD components of related indicators and use this identified relationship for prediction. This new prediction approach can be applied to any system of nonlinear and/or nonstationary variables.

3.3.1 Intrinsic Time-Scale Decomposition

Given a nonlinear time series signal, X_t , where $t \in (1, 2, \dots, T)$, a general representation of a dynamical system of X_t can be represented using Volterra expansion [97] as

$$X_t = \delta_t + \sum_{i=0}^{\infty} \omega_i \eta_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \omega_{ij} \eta_{t-i} \eta_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \omega_{ijk} \eta_{t-i} \eta_{t-j} \eta_{t-k} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \omega_{ijkl} \eta_{t-i} \eta_{t-j} \eta_{t-k} \eta_{t-l} + \dots \quad (3.33)$$

where δ_t is the deterministic component, $\eta_t \sim \text{IID}$, and $\{\omega_i, \omega_{ij}, \omega_{ijk}, \omega_{ijkl}, \dots\}$ is a set of coefficients. Eq. (3.33) is nonlinear if it has nonzero coefficients $\{\omega_{ij}, \omega_{ijk}, \omega_{ijkl}, \dots\}$ on the higher-order terms. The Volterra expansion in Eq. (3.33) can be transformed into an autoregressive representation [98] as

$$X_t = f(X_{t-1}, X_{t-2}, \dots) + \eta_t \quad (3.34)$$

where $f(X_{t-1}, X_{t-2}, \dots)$ is some nonlinear function of the past values of X . The difficulty in estimating Eq. (3.34) is that the functional form $f(\cdot)$ is unknown when working with real-world data. A number of procedures developed to determine the form of nonlinearity have been reported in the literature. However, no set of tests can actually discern the correct form of nonlinearity. Rather, the tests can only suggest the form of the nonlinearity.

Intrinsic time-scale decomposition (ITD) is a recently developed nonparametric algorithm aiming to enhance the analysis of nonlinear and nonstationary signals [99]. The advantage of ITD is that

it does not require the parametric functional form of the original series. The ITD method represents a nonlinear, nonstationary signal as a summation of components called proper rotation components and a monotonic trend. ITD decomposes the signal into a sequence of proper rotations of successively decreasing instantaneous frequency at each subsequent level of the decomposition as

$$X_t = HX_t + LX_t = HX_t + (H + L)LX_t = (H(1 + L) + L^2)X_t = (H \sum_{k=0}^{a-1} L^k + L^a)X_t \quad (3.35)$$

where H is defined as a proper-rotation-extracting operator and L is defined as a baseline-extracting operator. $HL^k X_t$ is the $(k+1)^{\text{st}}$ level proper rotation and $L^a X_t$ is the monotonic trend. The following ITD procedure is used to extract proper rotation components from a time series X_t :

- (1) Identify the location of all local extrema of X_t denoted as τ_k where $k \in \{1, 2, \dots\}$
- (2) Assume that X_t is available for $t \in [0, \tau_{k+2}]$ and define a baseline signal L_t using a piecewise linear baseline-extracting operator, L , between successive extrema as follows:

$$LX_t = L_t = L_k + \left(\frac{L_{k+1} - L_k}{X_{k+1} - X_k} \right) (X_t - X_k), \quad t \in (\tau_k, \tau_{k+1}] \quad (3.36)$$

where X_k and L_k denote $X(\tau_k)$ and $L(\tau_k)$ respectively, and

$$L_{k+1} = \alpha [X_k + \left(\frac{\tau_{k+1} - \tau_k}{\tau_{k+2} - \tau_k} \right) (X_{k+2} - X_k)] + (1 - \alpha) X_{k+1}, \quad 0 < \alpha < 1 \quad (3.37)$$

- (3) The first proper rotation can be extracted from X_t using a proper-rotation-extracting operator, H , as

$$HX_t \equiv (1 - L)X_t = H_t = X_t - L_t \quad (3.38)$$

(4) The steps (1) to (3) will perform iteratively, using the baseline signals as input, until the monotonic baseline signal is obtained. The following properties provide some key concepts of ITD:

Remark 1: The decomposition is nonlinear in the sense that the decomposition of two signals need not produce components equal to the sum of the components obtained from decomposing each signal individually, i.e. $S_t = X_t + W_t \Rightarrow HL^k S_t \neq HL^k X_t + HL^k W_t$

Remark 2: The edge effect of the ITD process is confined to the interval $[0, \tau_2]$ at the beginning and $[\tau_k, \tau_{k+2}]$ at the ending of each proper rotation and monotonic trend.

3.3.2 ITD-based Prediction Methodology

As mentioned in Section 3.3, the ITD-based prediction methodology takes advantage of the SRI methodology to identify the relationships of variables in the system. The ITD-based prediction methodology consists of the following three steps. The first step is an ITD step. In this step, an indicator is decomposed into a finite number of proper rotation components and the monotonic trend. The second step is the variable selection step. This step is used to select endogenous and exogenous variables for VECM. In the second step, the nonstationary condition and integration order for each component is identified using a unit root test. Since the endogenous variable is nonstationary, stationary ITD components are treated as exogenous variables. Other exogenous variables are identified using a weakly exogenous test. The long-run equilibrium relationships between endogenous variables are identified using a cointegration test. The third step is the prediction step. The VECM model of selected endogenous and exogenous variables in the second step is parameterized. Out-of-sample forecasting is achieved. A comparison of out-of-sample forecasting with other classical and advanced time series techniques is done in this step. The overall framework of the ITD-based prediction methodology for a system of automobile demand and economic indicators is shown in Figure 3.8.

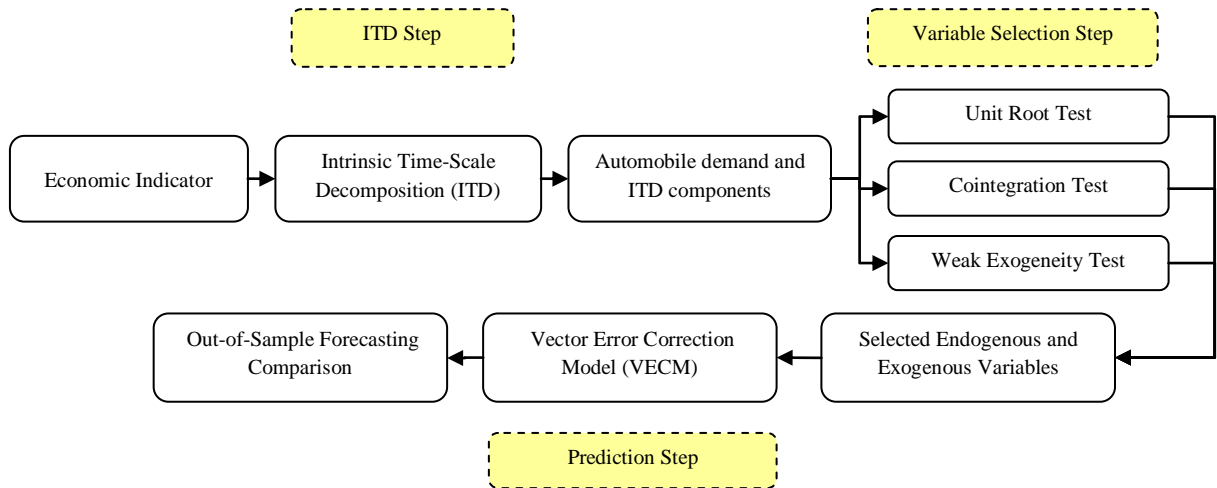


Figure 3.8: ITD-based Prediction Methodology Framework

CHAPTER IV

WILLINGNESS-TO-PAY PREDICTION BASED ON EMPIRICAL MODE DECOMPOSITION

Abstract: Long-term prediction of customer preferences is becoming essential for effective product portfolio design in broad industrial sectors such as automotive, aerospace, and consumer electronics, where typical concept-to-release times are long (24-60 months). However, nonlinear and nonstationary evolutions of customer preferences hinder their accurate prediction. The two-step prediction approach based on EMD described in Chapter 3 is proposed to forecast customer preferences over extended time-horizons. The advantage of EMD is that this method can be used to decompose a nonstationary time series into a finite number of components called intrinsic mode functions (IMFs). This property helps in isolating trend and noise components (detrending and denoising) in a nonstationary process. However, the presence of edge artifact limits the use of EMD for predictions. A key aspect of this two-step approach is that it takes advantage of the linear phase property of the Hilbert-Huang transform (HHT) to address the edge artifact limits, thus extending EMD for long-term prediction applications. The empirical results suggest that the accuracy of EMD-based prediction is significantly improved in terms of RMSE (36%) and R^2 (30%) compared to classical and advanced time series techniques.

4.1 Introduction

During the last decade, U.S. auto manufacturers have lost significant market share to foreign competitors. Train [100] investigated the reasons behind this loss of market share using a model of the influence of vehicle attributes on customer preference and found that tracking customer preferences is essential to the effective design of product portfolios. Customer preferences are often expressed in terms of willingness-to-pay (WTP) metrics [101], which refers to the maximum monetary amount a consumer is willing to pay for a good or service. Marketing researchers have used choice models to capture WTP and predict market share under certain varying market conditions [102, 103]. However, many of the existing choice models of customer preferences [104, 105] assume time invariant behavior of the customer. They seldom consider the evolving nature of customer choice over time. For complex product systems (e.g., automobiles) where the concept-to-release time takes approximately 24-60 months, customer preferences evolving over time may have a large impact on market performance. Here, WTP can be estimated for each of the attributes of the product. Examples of the major attributes of an automobile product are brand, fuel economy, durability, safety, workmanship, and initial reliability. Input from prospective customers elicited from on-line portals, social networks, and blogs, as well as product clinics is being considered for the estimation of WTP [106].

The WTP data in this study comprise a set of daily estimates of three fuel economy WTP attributes during a period of 24 months (731 data points). This dataset contains real WTP data capturing the trends of a typical automotive product from industry sources. However, the data have been masked for proprietary reasons. Fig. 4.1 shows the WTP for one attribute, fuel economy, and its autocorrelation. As shown in the figure, the series exhibits strong levels of nonstationarity. Since WTP evolves over time, one would expect that the design or estimate using past customer preferences (WTP) may perform sub-optimally in terms of market shares in the future. A solution to this issue might be to forecast the future values of WTP. However, the

typically long (24-60 month) concept-to-release time, coupled with the nonstationary characteristic of WTP pose significant challenges to long-term prediction. Very few efforts [107, 108] have addressed forecasting for long-term horizons. For stationary processes, unbiased long-term forecasting would be reduced to invariant first- and second moment estimation. In the case of nonstationary processes, long-term forecasts would be dominated by trend curves [109]. Currently, no simple approach or general theory exists for long-term forecasting under nonstationarity.

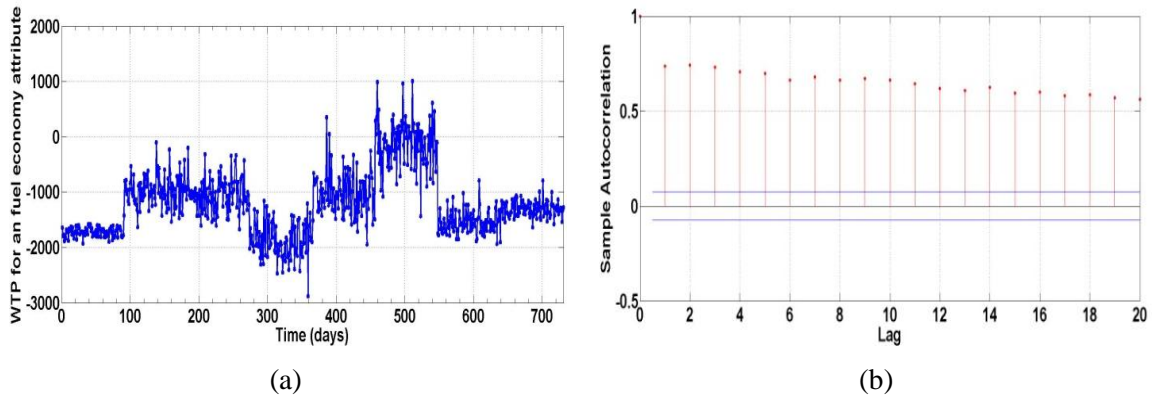


Figure 4.1: (a) Typical trend of WTP of a fuel economy attribute (b) Autocorrelation of WTP attribute

This study presents a new prediction approach based on the empirical mode decomposition (EMD) technique. EMD aims to decompose a nonstationary signal into a finite number of components called intrinsic mode functions (IMFs). EMD allows perfect reconstruction of the original signal using IMFs. This property leads to the application of determining a trend from an original series. Since WTP exhibits nonstationary characteristic, the EMD technique can be extended for long-term forecasting of WTP. This investigation indicates that EMD prediction can improve the forecasting of WTP over a 12-15-week horizon by about 30% (R^2) compared to other methods tested. The organization of this chapter is as follows: Section 4.2 presents the implementation details and results of using the EMD-based prediction methodology described in Chapter 3. The conclusion and suggested future work are presented in the last section of this chapter.

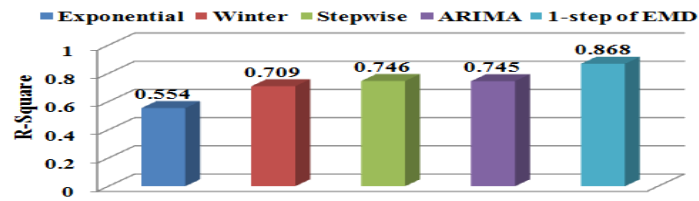
4.2 Implementation Details and Results

4.2.1 The First Step

The two-step algorithm approach was applied to the WTP for the fuel economy attribute. The first step of the two-step algorithm approach provides the long-term trend prediction of WTP which can be used to solve the edge effects problem of EMD. Exponential smoothing, the winter additive method, STEPARG, and ARIMA were selected to present the effectiveness of the EMD-based prediction model and to compare the prediction accuracy in terms of RMSE and R^2 . Prediction comparison (the first step of two-step algorithm approach) for the WTP attribute is summarized in Table 4.1. Comparing the five prediction models in Table 4.1, the first step of the EMD-based prediction is the best model to train (1-step ahead) the original WTP attribute in terms of RMSE (267.25) and R^2 (0.868), followed by STEPARG, ARIMA, winter additive, and exponential smoothing. From Table 4.1, applying STEPARG and ARIMA models directly to the original WTP data yields similar prediction accuracy ($R^2 \approx 75\%$). Using STEPARG with trend IMFs and residue in the first step of the two-step algorithm approach yields a significant improvement in terms of RMSE ($\approx 25\%$) and R^2 ($\approx 12\%$), compared to STEPARG (the second best model).

Table 4.1: Prediction Comparison (the first step of two-step approach)

Model	RMSE	R^2
Exponential Smoothing	458.18	0.554
Winter Additive Method	382.11	0.709
Stepwise Autoregressive	356.22	0.746
ARIMA	356.63	0.745
1-step of EMD-based prediction	267.25	0.868



4.2.2 The Second Step

The prediction results in the first step are then used to solve the edge effects problem of EMD. The results in the second step were obtained by applying Eq. (3.12) with all new IMFs and the residue. The second step of the two-step algorithm approach provides a long-term trend prediction as well as a short-term prediction of WTP. Prediction comparison (the second step of the two-step algorithm approach) for the original WTP attribute is summarized in Table 4.2. Comparing six prediction models, the second step of the EMD-based prediction is the best model to train (1-step ahead) the original WTP attribute in terms of RMSE (191.81) and R^2 (0.932). As seen in Table 4.2, applying the second step of EMD-based prediction improves the prediction accuracy in terms of RMSE (46%) and R^2 ($\approx 20\%$). The improvement in prediction accuracy results from both short-term IMF components (detrended components) and instantaneous phase predictions. Table 4.3 shows RMSE and R^2 with 1-5 (short-term), 6-10 (mid-term), and 11-15 (long-term) steps with validation data. The RMSE and R^2 values in Table 4.3 were achieved from applying different models with training data for three WTP attributes, and then validating these models with validation data. The results show that EMD-based prediction (the second step) is the best model for all short-, mid- and long-term predictions in terms of RMSE and R^2 . For long-term prediction, the two-step algorithm approach yields a significant improvement in terms of RMSE (36%) and R^2 (30%), compared to the average RMSE and R^2 of the other four models in Table 4.3.

Table 4.2: Prediction Comparison (the second step of the two-step algorithm approach)

Model	RMSE	R^2
Exponential Smoothing	458.18	0.554
Winter Additive Method	382.11	0.709
Stepwise Autoregressive	356.22	0.746
ARIMA	356.63	0.745
1-step of EMD-based prediction	267.25	0.868
2-step of EMD-based prediction	191.81	0.932

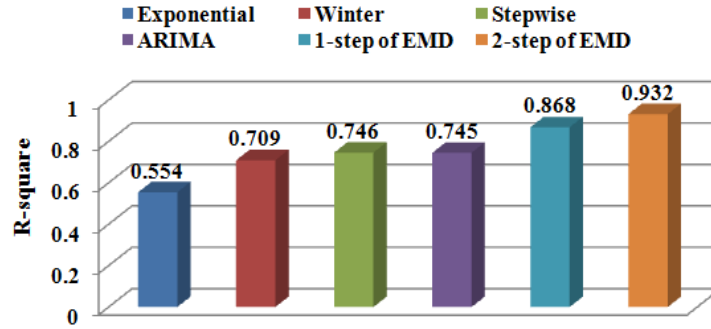
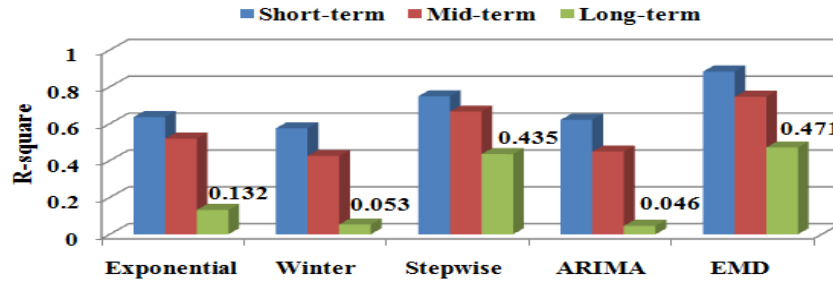


Table 4.3: Comparison of prediction accuracy with the validation data set

Model	Short-term		Mid-term		Long-term	
	RMSE	R ²	RMSE	R ²	RMSE	R ²
Exponential Smoothing	370.450	0.636	445.074	0.519	750.362	0.132
Winter Additive Method	400.253	0.575	486.639	0.425	818.762	0.053
Stepwise Autoregressive	307.275	0.749	370.654	0.666	488.422	0.435
ARIMA	377.757	0.621	476.234	0.449	824.921	0.046
EMD-based prediction	210.135	0.883	323.381	0.746	456.872	0.471



4.3 Conclusions and Suggested Future Work

Long-term prediction of customer preferences is becoming essential for effective product portfolio design in broad industrial sectors where typical concept-to-release times are long (24-60 months). However, the nonlinear and nonstationary evolution of customer preferences hinders their accurate prediction. Empirical mode decomposition (EMD), as a nonlinear and nonstationary signal processing method, offers the advantage of capturing nonstationarity in the signal by determining trend and noise components. However, the edge effects problem limits the use of EMD for prediction applications. In this study, a new two-step prediction approach based

on EMD was presented to facilitate long-term prediction of the highly nonstationary evolution of customer preferences. The first step procedure was used for predicting the long-term trend component of the customer preference series (expressed in terms of WTPs). In this step, EMD decomposed the original series into multiple IMFs (about 11 in most cases investigated) and a residue. The long-term trend was found to be embedded in IMFs 8 through 11, and the residue. The empirical results have shown that the first step of the two-step approach yields a notable improvement in terms of RMSE (25%) and R^2 (12%), compared to STEPAR. The first step of the two-step algorithm prediction result does not just provide better prediction accuracy, but it can also be used to reduce the edge effects problem of EMD.

In the second step of the two-step algorithm, EMD was used to decompose an extended series obtained from concatenating the original series with prediction result from the first step. In this step, EMD provided new IMFs and residue without the edge effects problem. Linear phase extension was then applied to all new IMFs and the residue. The empirical results show that applying the second step of EMD improves the accuracy of predicting WTP over extended time horizons both in terms of RMSE (46%) and R^2 ($\approx 20\%$). The improvement of prediction accuracy results from both short-term IMF components (detrended components) and instantaneous phase predictions. Overall, the two-step prediction approach based on EMD appears to be suited well for the prediction of nonstationary processes. However, in this study, the underlying process and related indicators of WTP were assumed to be unknown. In principle, the prediction should not be done with historical data (autoregressive components) alone. Exogenous variables that affect the signals of interest are absolutely essential. To improve the prediction accuracy, future work in prediction based on EMD should involve incorporating exogenous variables into the model.

CHAPTER V

MULTI-STEP AUTOMOTIVE SALES FORECASTING BASED ON STRUCTURAL RELATIONSHIP IDENTIFICATION

Abstract: Forecasting sales and demand over a 6-24-month horizon is crucial for planning the production processes of automotive and other complex product industries (e.g., electronics and heavy equipment) where typical concept-to-release times are 12-60 months. However, nonlinear and nonstationary evolution and dependencies with diverse macroeconomic variables hinder accurate long-term prediction of the future of automotive sales. In this chapter, a structural relationship identification (SRI) methodology, described in Chapter 3, that uses a battery of statistical unit root, weak exogeneity, Granger-causality, and cointegration tests, is proposed to identify the dynamic couplings among automobile sales and economic indicators. A vector error correction model (VECM) of multi-segment automobile sales was estimated based on impulse response functions to quantify the long-term impact of these economic indicators on sales. Comparisons of prediction accuracy demonstrate that the VECM model outperforms other classical and advanced time-series techniques. The empirical results suggest that VECM can significantly improve the accuracy of 12-month ahead automotive sales predictions in terms of RMSE (42.73%) and MAPE (42.25%), compared to the classical time series techniques.

5.1 Introduction

Competition among automobile manufacturers has been growing rapidly since the U.S. economic crisis in 2008 [110]. Manufacturers are forced to devise various strategies to help them compete in the industry. In order to effectively manage resources and maximizing revenue, it is imperative that product portfolios be designed and positioned suitably and that production activities (e.g. production and operations planning, inventory and material management, etc.) also be planned correctly. Because of long product development cycles, effective planning for production of automobiles requires accurate forecasting of long-term (6-24 months) sales and demand. Long-term forecasting serves as an input to many business and operating decisions which affect profitability. In production planning processes, long-term forecasting is used to determine an appropriate level of manpower for production planning and also serves as an input for business planning, such as planning for expansion or contraction of existing production units. Errors in demand forecasts have often led to enormous costs and loss of revenues. For example, inaccurate forecasts of future demand have led to highly suboptimal levels of production workforce [111]. The consequent workload imbalances can increase the costs of hiring, firing, and overtime labor activities. In many cases where the supply of materials is outsourced, inaccurate demand forecasts can also lead to a shortage of supplies for production. Sales forecasting [112-114] is considered a realistic assessment of expected future demand [115].

Automobile sales forecasting has received notable attention in the past 30 years [116-123]. For example, Berkovec [116] modeled the aggregate automobile market demand as the sum of individual consumer demand using household attributes and vehicle ownership levels. Greenspan and Cohen [117] predicted aggregate sales of new motor vehicles based on quantifying vehicle stock and scrappage rates. Franses extended the deterministic Gompertz model [124, 125] to predict Dutch new car sales. Armstrong *et al.* [121] developed intentions-based forecasting methods to predict French and US automobile sales. Kunhui *et al.* [123] predicted Chinese

automobile sales using support vector regression (SVR). Hulsmann *et al.* [122] applied data mining algorithms to model German and US automobile markets. Based on the literature, most proposed automobile sales forecasting models are econometric approaches imposing a certain structure of economic theory on the data. There is no guarantee that the underlying processes of automobile sales will follow these structures.

Automobile sales are known to be influenced by factors such as advertising, sales promotions, retail prices, and technological sophistication [126]. Considering the long concept-to-release time (12-60 months) of automobile products, effective production planning requires a forecast of sales in a long-term horizon. In a competitive market, as in the case of the automobile market, advertising and sales promotions tend to have substantial effects; however, these effects on sales are rarely persistent [127-130]. Factors reflecting customer preferences in the automobile market, such as aesthetics [131], brand, form, and function, [132], can also influence automobile sales at segment or trim levels, but not in determining changes in automobile sales at aggregate levels.

Recently, a relatively small body of literature has provided a set of economic indicators that relate to sales [133, 134]. Mian and Sufi [134] revealed the similarities among the evolution patterns of automobile sales, unemployment rates, and new housing building permits (Housing Starts). McManus [133] presented a link between gasoline prices and vehicle sales. However very little effort has been made to address automobile sales prediction using these economic indicators [135-137]. Wang *et al.* [137] used an adaptive network-based fuzzy inference system to estimate new automobile sales in Taiwan with economic indicators. Shahabuddin [136] modeled vehicle sales using indicators such as durable personal consumption and durable industrial demand. Brühl *et al.*'s [135] data-driven model for the German automobile market relates various economic indicators such as gross domestic product (GDP), the Consumer Price Index (CPI), interest rates, the unemployment rate, and gas prices with automobile sales. The results using the empirical models suggest that these indicators have a significant effect on automobile sales. However, these

models are applicable to forecasting horizons that are much shorter (1-4 months) than those needed for effective production planning for automobile sales (6-24 months).

Apart from its multivariate nature, the nonstationary evolution of many of these variables presents additional challenges to automobile sales prediction for a long-term horizon. For stationary processes, unbiased long-term forecasting would be reduced to invariant first and second moments. However, in nonstationary processes, the evolution of the systems can deviate permanently from previous levels, which makes long-term forecasting almost impossible unless the underlying structures or processes of the systems are known. In the area of econometric time series research, some developments in multivariate time-series techniques have been specifically designed to quantify the structural relationship of variables in the system. These models are the vector autoregressive (VAR) and vector error correction models (VECM) [75, 76]. They have been broadly recognized as powerful, theory-driven models that can be used to describe the long-run dynamic behavior of multivariate time series. Especially in the case of nonstationary variables with cointegration, the VECM provides the best persistent modeling for addressing the problem of long-run market-response identification by estimating the cointegration vectors of the endogenous variables. However, with this type of model, the identification of the structural relationship of the variables is often problematic. For example, the large number of endogenous and exogenous variables may cause an over-parameterization issue leading to wrong estimations of all model coefficients. Model misspecification in VAR and VECM may also be due to including irrelevant endogenous variables in the models.

This study presents a structural relationship identification (SRI) methodology to address these pertinent issues in VAR and VECM for automobile sales prediction. The hypothesis is that a causal relationship and/or cointegration exist in our dynamical system of multi-segment automobile sales and economic indicators. In the prediction of the sales of multi-automobile segments, the interaction effect among segment-wise sales can be significant due to the common

influence of multiple economic indicators. In such cases, the present VAR and VECM structures are expected to perform better than other traditional time series models because of their inherent advantages [96, 138]. For example, VAR and VECM can be used to capture the linear interdependencies of a multivariate time series system and to provide evidence of the dynamic behavior and structural relationships among variables [139]. These models also allow a feedback relationship, which can be used to capture more complex interactions among variables [140].

The methodology used in this study includes a battery of methods to test for a unit root, weak exogeneity, Granger causality, and cointegration. The unit root test identifies the stationarity of each variable in the model. A sequential reduction method for weak exogeneity addresses an over-parameterization problem and also identifies the reasonable structural relationships of endogenous variables. Granger causality identifies causal relationships for the variable selection process. Because of the nonstationary characteristic of the variables, a cointegration test is used to identify the existence of long-run equilibrium relationships among selected endogenous variables. The empirical results indicate that long-run equilibrium relationships among variables in dynamic systems do exist. Hence, in this study, VECM is derived and implemented for predicting sales in two different segments and is compared with other existing methods. The long-term impact of each economic indicator on automobile sales are also discussed. The organization of the remainder of this chapter is as follows: Section 5.2 provides the description of each variable in the multi-segment automobile sales model. The implementation details and empirical results are given in Section 5.3. Conclusions and further discussion are presented in Section 5.4.

5.2 Data

The main time series considered in this study is the sequence of monthly new automobile sales in the U.S., as reported in industry sources, during the period Jan 1975 – Dec 2010. The dataset consists of new automobile sales in two different segments (large- and small-vehicle segments) and selected economic indicators. The large-vehicle segment is the aggregate number of automobile sales in heavy and medium truck segments. The small-vehicle segment is the aggregate number of automobile sales in sedan and SUV segments. Although large- and small-vehicle segments seem to have different structures that control the underlying dynamics in the short term, we can observe a similarity in long term trends from both segments as shown in Fig. 5.1.

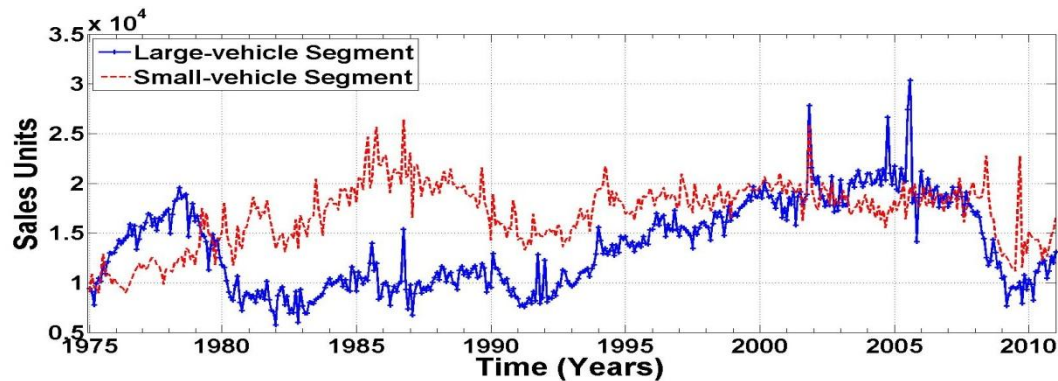


Figure 5.1: Sales of Large- and Small-Vehicle Segments

Considering sales of automobiles in both segments, the hypothesis is that both segments have a long-run equilibrium relationship which can lead to better estimation of future sales for long-term horizons. The remaining variables in this dataset are selected economic indicators. As discussed in the introduction, the selection of economic indicators is intended to improve the prediction of automobile sales, leading to an improvement in production planning efficiency for long-term horizons. A sufficient number of economic indicators must be selected to reveal a structural

relationship between them and automobile sales. The following properties are considered for the selection of economic indicators in this study:

- i. Indicators that characterize changes in the price paid by automobile consumers.
- ii. Indicators that influence the demand behavior in large- and small-vehicle segments of the automobile industry.
- iii. Indicators that represent the national economy and changes in the economic cycle.

In addition to the properties described above, all aspects of modeling issues, such as over-parameterization, multicollinearity, redundancy, and model specification, are considered in selecting the variables. From the literature and from preliminary tests as well as guidance from industry experts, four monthly economic indicators are selected (See Fig. 5.2): Consumer Price Index (CPI), Unemployment Rate, Gas Prices and Housing Starts.

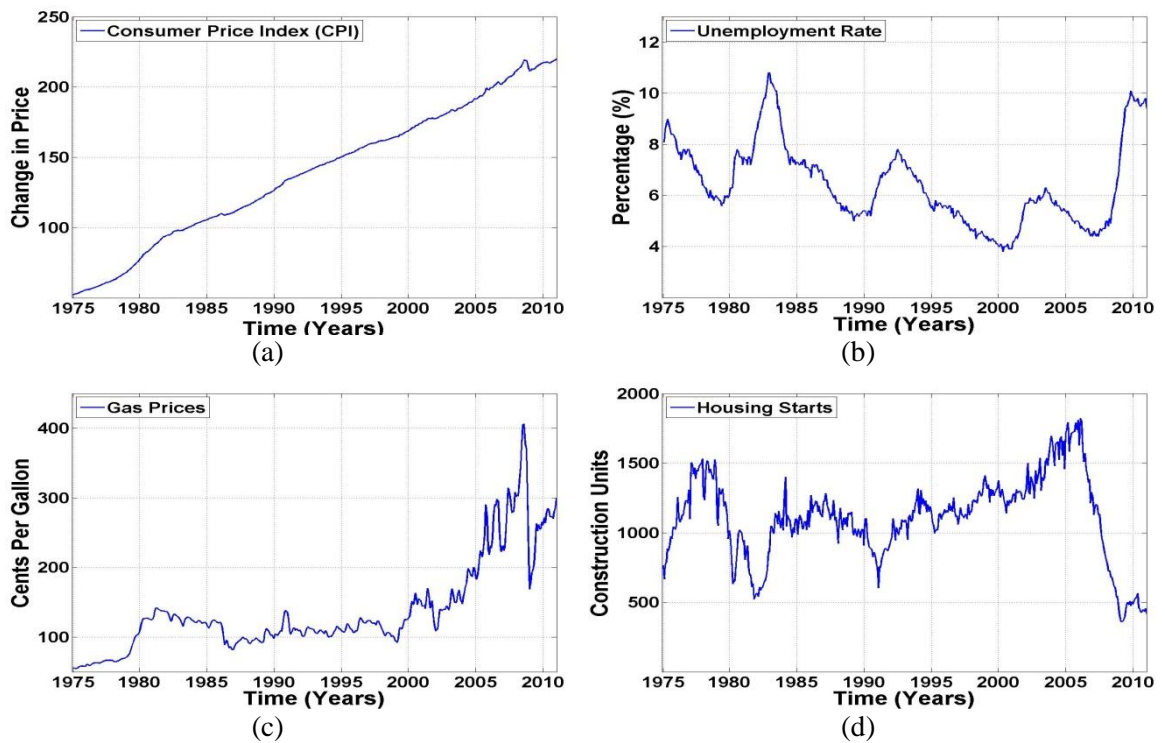


Figure 5.2: Economic Indicators: (a) Consumer Price Index (CPI), (b) Unemployment Rate, (c) Gas Prices and (d) Housing Starts

These four indicators are hypothesized to be sufficient for modeling automobile sales for long-term horizons, considering the selection criteria. Although some of these indicators can be influenced by common underlying structures resulting in their being significantly correlated, each of these indicators represents different characteristic properties that should be addressed. The structural identification methodology presented in this paper is used to identify the structural relationship between automobile sales and these economic indicators.

As shown in Fig. 5.2, most of the selected economic indicators exhibit nonlinear trends or patterns. Gas prices tend to grow exponentially over time due to increases in the scarcity of resources. Unemployment rates as well as housing starts seem to have a cyclical pattern resulting from economic up- and downturns; however, housing starts tend to have a greater level of short-term fluctuation. Only the CPI shows a linear trend over time. Data transformation is required to satisfy the linearity assumption underlying VAR and VECM in order to model a system of automobile sales and economic indicators. In this paper, a logarithmic transformation is selected for all variables. By taking the log transformation, not only can data be converted from nonlinear to linear patterns, but the variability of variables can also be stabilized. Table 5.1 summarizes automobile sales data and the selected economic indicators for the purpose of prediction. To preserve data for out-of-sample forecast evaluation, the model is identified using a truncated sample from Jan 1975 to Dec 2009, and the remaining data is used for validation. For reasons of confidentiality, these variables are masked in the analysis. The sales variables and economic indicators in Table 5.1 are represented by Y_i 's and X_i 's respectively. The order of these variables in our analysis and results also differs from the order in Table 5.1.

Table 5.1: Summary of variables

Variable	Name	Source	Description
Y_1	Small-Vehicle Segment Sales	Industry	Aggregate small-vehicle segment sales of sedan & SUV models.
Y_2	Large-Vehicle Segment Sales	Industry	Aggregate large-vehicle segment sales of heavy & medium-sized trucks.
X_1	Consumer Price Index	BLS**	Monthly data on changes in the prices paid by urban consumers for a representative basket of goods and services (Base Period: 1983=100).
X_2	Unemployment rate	BLS**	National unemployment rate (16 years or over).
X_3	Gas Prices	EIA***	Monthly U.S. retail prices of regular gasoline, all formulations.
X_4	Housing Starts	Census Bureau and HUD*	Monthly data on the construction of private residential structures, such as single-family homes and apartment buildings.

* Department of Housing and Urban Development, ** Bureau of Labor Statistics, *** U.S. Energy Information Administration

5.3 Implementation Details and Results

In this section, implementation details and empirical results of the structural identification methodology (unit root, weak exogeneity, cointegration, and Granger causality) described in Chapter 3 are presented. The unit root test results indicate that automobile sales and economic indicators are nonstationary variables. Based on the results of the weak exogeneity tests, two economic indicators (X_1 and X_3) are treated as exogenous variables. The cointegration rank test results indicate that automobile sales and two economic indicators (X_2 and X_4) are cointegrated. Finally, the VECM and VECMX models of automobile sales and economic indicators are estimated, and an out-of-sample forecasting comparison of the VECMX and classical time series techniques shows that VECMX significantly improves the accuracy of long-term prediction of automobile sales.

5.3.1 Unit Root

In order to conduct the ADF unit root test, it is important to use the correct number of lags in the equation. The optimal lag length of the original variables for the ADF test was individually selected based on the results of general-to-specific methodology (t -test). The methodology starts with a lag length of $p^*=6$ and then pares down the model by the usual t -test. If the t -statistic on lag p^* is insignificant at some specified critical value, then the regression is re-estimated using a lag length of p^*-1 . The process is repeated until the last lag is significantly different from zero. In this autoregressive case, this procedure yields the true lag length with an asymptotic probability of unity, provided the initial choice of lag length includes the true length. Another technique for selecting optimal lag length for the ADF test is to use information criteria such as the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC), whose formulas are specified respectively as

$$AIC = -2\ln(L)/T + 2n/T \quad (5.1)$$

$$SBC = -2\ln(L)/T + n\ln(T)/T \quad (5.2)$$

where $\ln(L) = -(T/2)\ln(2\pi) - (T/2)\ln(\sigma^2) - (1/2\sigma^2)(SSR)$, SSR = sum of squared residuals, σ^2 = variance of the residuals, n = number of parameters estimated, and T = number of usable observations.

To avoid comparing the models over different sample periods, the number of usable observations was fixed at 414 observations (overall observations – maximum lag length = 420 – 6 = 414 observations). The information criteria (AIC and BIC) for each p^{th} order autoregressive process is reported in Table A1 (Appendix A).

The Akaike information criterion (AIC) results are mostly consistent with the optimal lag length selection using a general-to-specific methodology (t -test) for all variables (in logarithm form), except $\log(Y_2)$ and $\log(X_1)$. The AIC for the $\log(Y_2)$ model selects a lag length of 6 and the SBC selects a lag length of 4. Since the SBC tends to select a more parsimonious model and the t -test

also selects lag 4, a lag length of 4 was selected as the optimal lag for the $\log(Y_2)$ variable. For the $\log(X_1)$ variable, the t -test selects a lag length of 3, the SBC selects a lag length of 1, and the AIC selects a lag length of 4. Nevertheless, in this case, the null hypothesis of the ADF test cannot be rejected for all lags (1, 3 and 4). The optimal lag length of 3 was selected to report the test statistics for the $\log(X_1)$ variable. In the case of differenced variables, the results of the ADF test are that the null hypothesis was rejected for lags 0-6 of all variables. The optimal lag length of each differenced variable was selected to report the ADF test results as shown in Table 5.2.

Table 5.2: ADF Unit Root Test of Original and Differenced Variables

Variables	Optimal Lag Length	Unit Root with Drift		Unit Root with Drift and Deterministic Time Trend	
		$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=1}^{p^*} \delta_i \Delta y_{t-i} + \varepsilon_t$		$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p^*} \delta_i \Delta y_{t-i} + \varepsilon_t$	
Original Variables	p^*	τ_μ	$Pr < \tau_\mu$	τ_τ	$Pr < \tau_\tau$
$\log(Y_1)$	5	-1.59	0.4889	-1.25	0.8978
$\log(Y_2)$	4	-3.18	0.0221	-2.74	0.2209
$\log(X_1)$	3	-1.61	0.4761	-1.47	0.8390
$\log(X_2)$	6	-3.19	0.0215	-2.43	0.3646
$\log(X_3)$	5	-1.40	0.5836	-2.44	0.4681
$\log(X_4)$	6	-2.68	0.0792	-2.49	0.3341
First Differenced Variables		τ_μ	$Pr < \tau_\mu$	τ_τ	$Pr < \tau_\tau$
$\Delta \log(Y_1)$	4	-12.64	<0.0001	-	-
$\Delta \log(Y_2)$	3	-15.24	<0.0001	-	-
$\Delta \log(X_1)$	1	-24.53	<0.0001	-	-
$\Delta \log(X_2)$	6	-3.82	0.0031	-	-
$\Delta \log(X_3)$	4	-5.14	<0.0001	-	-
$\Delta \log(X_4)$	5	-4.72	0.0002	-	-

The ADF test for unit root with drift and deterministic time trend on the first differenced variables are not computed. Test critical values for the unit root with drift model: 1% (-3.446), 5% (-2.868) and 10% (-2.570). Test critical values for the unit root with drift and linear time trend model: 1% (-3.980), 5% (-3.421) and 10% (-3.133). Lag length selection criteria: AIC, BIC and t-test.

Table 5.2 lists the ADF test statistics of the original and differenced variables for the unit root with drift and the unit root with drift and deterministic time trend models. The test results show

that the null hypothesis of unit root cannot be rejected for all original variables at the 1% significance level but is rejected at the 5% significance level for $\log(Y_2)$ and $\log(X_2)$ in the unit root with drift model. However, the null hypothesis for a unit root cannot be rejected for these variables in the unit root with drift and deterministic time trend model. These results indicate that not all variables are linear-trend stationary variables. For the first differenced variables, the ADF test statistics reject the null hypothesis of a unit root with drift for all differenced variables.

The χ^2 -value of the univariate G-test of each original variable with $p=1$ and $q=2$ to 8 is reported in Table A2 (Appendix A). The empirical results show that the null hypothesis of stationarity is rejected at the 10% significance level for all original variables at all levels of $G(p,q)$, except $\log(Y_1)$ and $\log(X_1)$ at the $G(1,2)$ level. However, for the first differenced variables (See Table A3, Appendix A), the null hypothesis of stationarity cannot be rejected for these variables at the $G(1,2)$ level. Based on the empirical results of two unit root tests, all original variables used in the model are treated as nonstationary $I(1)$ series in the subsequent analysis.

5.3.2 Weak Exogeneity

A sequential reduction method for the weak exogeneity test begins with the vector of the log transformation of the six variables (Y_i 's & X_i 's) in Table 5.1. For a system with six variables, there can exist at most a cointegrating vector with rank five. The null hypothesis of $H_0: \alpha_2 = 0$ for each candidate exogenous variable is tested with the cointegrating rank unrestricted ($r = 5$). The results in Table 5.3 show that the null hypothesis of weak exogeneity cannot be rejected for $\log(X_3)$. In contrast, the null hypothesis of weak exogeneity for the other variables ($\log(Y_1), \log(Y_2), \log(X_1), \log(X_2)$ and $\log(X_4)$) is strongly rejected at the less than 1% significance level.

Greenslade *et al.* [89] suggested retesting the rejecting weakly exogenous variables with the non-rejecting weakly exogenous variables to avoid sensitivity on the model specification. The results

of retesting the weak exogeneity variables are given in Tables 5.4 and 5.5. The test results continue to strongly reject the null hypothesis of weak exogeneity of $\log(Y_1)$, $\log(Y_2)$, $\log(X_2)$ and $\log(X_4)$ at less than a 1% significance level. The null hypothesis of weak exogeneity cannot be rejected in the second test in the case of $\log(X_1)$; therefore $\log(X_1)$ and $\log(X_3)$ are treated as exogenous variables in subsequent analysis.

Table 5.3: Weak Exogeneity Tests (all variables)

Variable	$\chi^2(5)$	P-value
$\log(Y_1)$	116.89	<0.0001
$\log(Y_2)$	79.27	<0.0001
$\log(X_1)$	23.08	0.0003
$\log(X_2)$	151.41	<0.0001
$\log(X_3)$	3.67	0.5973
$\log(X_4)$	87.63	<0.0001

Table 5.4: Weak Exogeneity Tests (Re-test(1))

Variable	$\chi^2(4)$	P-value
$\log(Y_1)$	111.78	<0.0001
$\log(Y_2)$	75.30	<0.0001
$\log(X_1)$	5.44	0.2450
$\log(X_2)$	126.94	<0.0001
$\log(X_4)$	80.89	<0.0001

Table 5.5: Weak Exogeneity Tests (Re-test(2))

Variable	$\chi^2(3)$	P-value
$\log(Y_1)$	53.01	<0.0001
$\log(Y_2)$	50.98	<0.0001
$\log(X_2)$	127.85	<0.0001
$\log(X_4)$	36.25	<0.0001

5.3.3 Cointegration Rank Test

The existence of long-run relationships among $\log(Y_1)$, $\log(Y_2)$, $\log(X_2)$ and $\log(X_4)$ was investigated; Table 5.6 reports the test statistics and the corresponding asymptotic critical values at the 5% and 10% significance levels. Based on both trace and maximum eigenvalue statistics, the null hypothesis of no cointegration is rejected. The null hypothesis of one cointegrating vector cannot be rejected at both 5% and 10% significance levels as shown in Table 5.6. The test results indicate one potential cointegrating vector among the four variables. Table 5.7 presents the estimated long-run parameter (β) and the adjustment coefficient (α) with $\log(Y_1)$ as a normalized variable. Therefore, the long-run relationship among the four variables is given by

$$\log(Y_1) = -2.4147\log(Y_2) + 5.5670\log(X_2) + 1.6327\log(X_4) \quad (5.3)$$

Table 5.6: Johansen's Cointegration Rank Tests

H_0 : Rank = r	H_1 : Rank > r	Trace Test			Maximum Eigenvalue Test		
		Trace Statistic	5%	10%	Max Statistic	5%	10%
0	0	71.776	39.71	36.58	58.203	23.80	21.58
1	1	13.573	24.08	21.58	8.018	17.89	15.59

Table 5.7: Long-Run parameter (β) and adjustment coefficient (α)

	β	α
$\log(Y_1)$	1	0.00077
$\log(Y_2)$	2.4147	-0.00270
$\log(X_2)$	-5.5670	0.00051
$\log(X_4)$	-1.6327	-0.00152

Weak exogeneity and cointegration test results show that both segments of automobile sales ($\log(Y_1)$ and $\log(Y_2)$) and two hypothesized economic indicators ($\log(X_2)$ and $\log(X_4)$) are I(1) and cointegrated of order 1. The underlying processes of these variables are random in the short term but tend to move together over a long-term horizon. As discussed in the methodology

section, the VECM is the most suitable model to explain the structural relationship of this system because it imposes cointegrating restrictions on endogenous variables.

5.3.4 Granger Causality Test

Although weak exogeneity and cointegration test results provide the information necessary for identifying the structural relationships of the system of multi-segment automobile sales and economic indicators, the Granger-causality test can still be used to confirm the exogeneity of variables. The Granger-causality test is used to test the causal relationships among six variables in order to confirm the exogeneity of $\log(X_1)$ and $\log(X_3)$. All variables in differenced form are built into a VAR system. In addition to AIC and SBC, the lag length criterion for the VAR specification also includes the corrected AIC (AICC), the Hannan-Quinn criterion (HQC), and the final prediction error criterion (FPEC). The results of lag length selection are given in Table A4 (Appendix A). All criteria select a lag length of four, except the SBC. Therefore, lag 4 is selected for the Granger-causality test.

To test the causal relationship, the chi-squared (Wald) test statistic was calculated for each equation in the VAR model. The results of Granger-causality tests (in Table 5.8) show that the null hypothesis cannot be rejected in the case of $\log(X_1)$ and $\log(X_3)$. The weak exogeneity and Granger-causality tests are consistent. These two economic indicators must be treated as exogenous variables.

Table 5.8: Granger Causality Tests

Lag Order	$\Delta \log(Y_1)$		$\Delta \log(Y_2)$		$\Delta \log(X_1)$		$\Delta \log(X_2)$		$\Delta \log(X_3)$		$\Delta \log(X_4)$	
	χ^2	P-value	χ^2	P-value	χ^2	P-value	χ^2	P-value	χ^2	P-value	χ^2	P-value
1	30.40	<0.0001	6.65	0.2482	6.27	0.2809	11.36	0.0447	1.39	0.9256	12.76	0.0257
2	44.73	<0.0001	24.00	0.0076	11.18	0.3435	55.04	<0.0001	11.74	0.3028	18.87	0.0419
3	55.38	<0.0001	34.54	0.0029	13.57	0.5581	53.08	<0.0001	14.67	0.4752	28.55	0.0184
4	60.41	<0.0001	41.66	0.0031	16.01	0.7162	54.95	<0.0001	20.27	0.4409	46.86	0.0006

χ^2 test statistics are presented for the null hypothesis that the column variable is influenced by itself and not by other variables. (Coefficients of all other variables, treated as independent variables, are not significantly different from zero)

5.3.5 Out-of-Sample Forecasting Performance Comparison

To evaluate the forecasting performance of the VECM model of automobile sales, the VECM model was developed with the sample period from Jan 1975 to Dec 2009. The 12-month period in 2010 was used to validate and compare the forecasting performance of the VECM model with other models. In this analysis, the VAR and VARX were selected to compare with the VECM and VECMX. The VAR and VARX admit the possibility of cointegration but do not impose cointegrating restrictions, as in the VECM and VECMX models.

The endogenous and exogenous variables in these models (VAR, VARX, VECM and VECMX) are explained by up to two lags ($p = s = 2$). Overall model criteria based on AIC, SBC, AICC, HQC and FEPC are presented in Table A5 (Appendix A). For forecasting accuracy comparison, the root mean square error (RMSE) and mean absolute percentage error (MAPE) for each model are computed based on the 12-step-ahead sales forecast as presented in Table 5.9.

Table 5.9: 12-step-ahead Forecasting Comparison (with trend, ADL and ARIMA(X) models)

MODEL	12-step-ahead Sales Forecast of $\log(Y_1)$ (out-of-sample)	
	RMSE	MAPE
Quadratic Trend Model*	0.6138	0.0649
IMA*	0.2848	0.0278
ARIMA*	0.2517	0.0235
ADL*	0.2659	0.0260
ARIMAX*	0.1866	0.0162
VAR	0.1719	0.0165
VARX	0.1737	0.0167
VECM	0.1416	0.0135
VECMX	0.1470	0.0139

*Modeling details are given in Table A6 (Appendix A)

Table 5.9 also provides the RMSE and MAPE values of five classical time series models (quadratic trend, integrated moving average (IMA), autoregressive integrated moving average (ARIMA), autoregressive distributed lags (ADL) and autoregressive integrated moving average

with exogenous variables (ARIMAX) models). The results show that the VECM model without exogenous variables outperforms all selected classical time series techniques and significantly improves prediction accuracy in terms of RMSE (42.73%) and MAPE (42.25%) for 12-step ahead prediction, compared to all selected classical time series techniques².

Fig. 5.3 shows the comparison of VAR, VARX, VECM and VECMX models for the forecasting of $\log(Y_1)$ in 2010. Although, the VECM cannot capture the short-term changes in $\log(Y_1)$ in 2010, it tends to be capable of capturing the long-term market-response in this segment of automobile sales. To validate the VECM for longer-term forecasting horizons, the average RMSE and MAPE were computed based on 18- to 24-step-ahead automobile sales forecasts of $\log(Y_1)$ as shown in Fig. 5.4³.

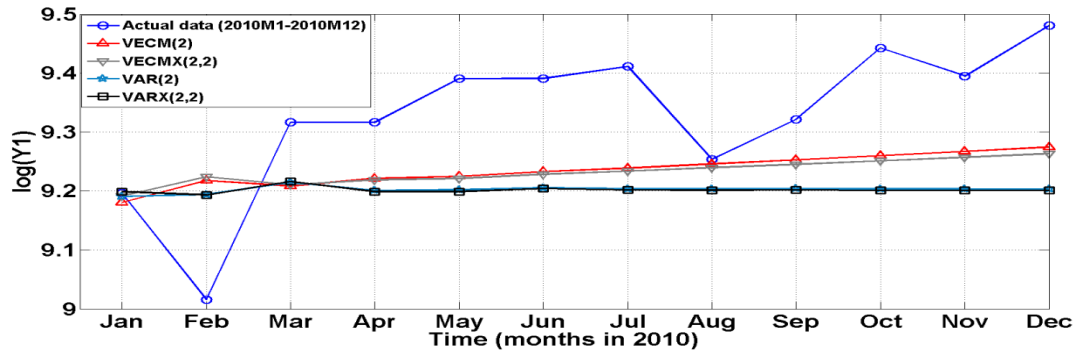


Figure 5.3: 12-step-ahead Forecasting Comparison (2010)

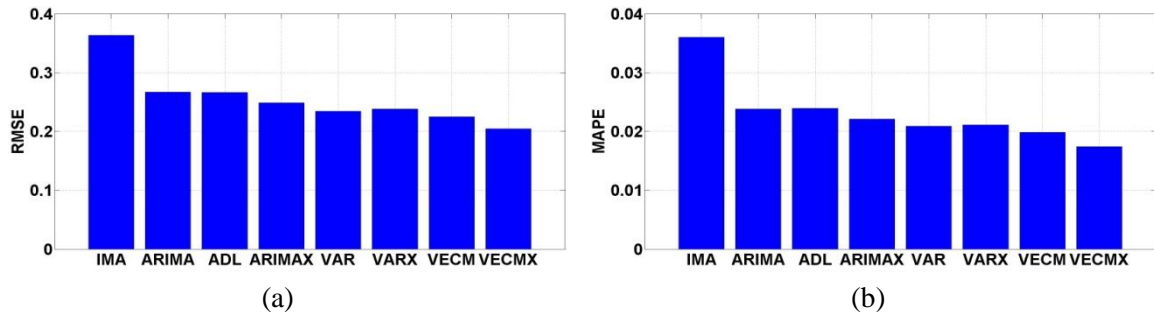


Figure 5.4: Model Selection Criteria Comparison (a) Average RMSE Comparison (b) Average MAPE Comparison

² To compute average RMSE and MAPE values of classical time series techniques, we exclude the worst model (Quadratic Trend Model) from our analysis.

³ Models in Fig. 5.5 were trained and validated based on an 18- to 24-step-ahead prediction.

The results show that VECMX outperforms all selected classical time series techniques⁴ in terms of RMSE and MAPE for longer-term forecasting horizons. Interestingly, adding exogenous variables (X_1 and X_3) does improve the forecasting performance for 18- to 24-step-ahead prediction, compared to the VECM without exogenous variables and all classical time series techniques.

5.3.6 Impulse Response Function

Impulse response functions derived from the VECM were used to trace the incremental effect of a one-unit shock in the endogenous variables on the future values of sales. In an automobile market, a one-unit shock in economic indicators, $\log(X_2)$ and $\log(X_4)$, seems to have a permanently negative impact on automobile sales ($\log(Y_1)$) as shown in Fig. 5.5(a) and 5.5(c). The increments of sales converge to some negative value. This convergence may imply that an increase in $\log(X_2)$ or $\log(X_4)$ can negatively influence $\log(Y_1)$ in the long run. Interestingly, an increase in a different segment of automobile sales, $\log(Y_2)$, also seems to have a persistently negative impact on $\log(Y_1)$, as shown in Fig. 5.5(b). In this case, the long-run effect has been identified, as an increase in Y_2 will decrease the sales in Y_1 . For Y_1 itself, an increase in $\log(Y_1)$ tends to have a positive impact from 1 to 12 months ahead, as shown in Fig. 5.5(d). However, the long-term trend of the impulse response seems to have a negative slope. From the impulse response functions of all endogenous variables, we can see that all variables seem to have a persistent effect on sales.

⁴ The quadratic trend model is excluded because of its worst-forecast performance.

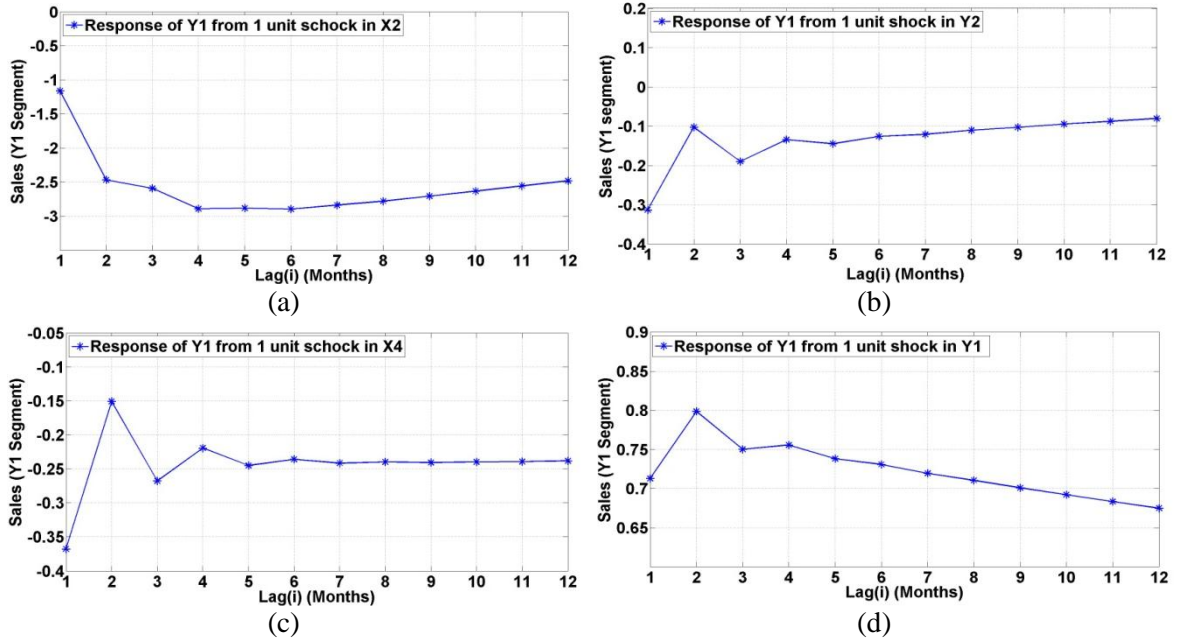


Figure 5.5: Impulse Response Functions from VECM: (a) Response of Y_1 from X_2 (b) Response of Y_1 from Y_2 (c) Response of Y_1 from X_4 and (d) Response of Y_1 from Y_1

5.4 Conclusions and Suggested Future Work

The VECM approach with SRI methodology to modeling automobile sales offers persistent modeling with two distinct advantages. First, it provides a clear and quantifiable method to estimate the long-run effect of economic indicators on sales. The second advantage is that it uses a feedback system approach to forecasting in estimating the market response to the economic indicators for which some advanced methods have been reported over the years. This study explores automobile sales modeling. Four economic indicators were selected to test the causal relationships and long-run equilibrium with sales of automobiles in two different segments. The unit root test results indicate the nonstationarity of all variables in the dataset. The test results also indicate that sales and selected economic indicators are first differenced stationary variables. From the four economic indicators, the conclusion based on the weak exogeneity and Granger-causality tests is that X_1 and X_3 are weakly exogenous to sales and other economic indicators. The

Johansen cointegration tests indicate a potential for a long-run equilibrium relationship among sales in two segments, X_2 and X_4 . The long run relationship among these variables can be quantified using the VECM approach. The response to the change in the one-unit shock of X_2 and X_4 to Y_1 implies a persistently negative impact on the sales. The increase in X_2 and X_4 values tends to have a negative impact on sales in the long run. The out-of-sample forecasting performance comparison of the VECM against the VAR(X) and classical time series techniques models shows that the VECM outperforms the VAR(X) and classical time series models in terms of RMSE and MAPE.

The weak exogeneity and Granger-causality test results indicate that X_1 and X_3 are weakly exogenous to sales and other economic indicators. However, the weakly exogenous relation may be due to the effect of absolute prices in the economic indicators. Automobile sales modeling using the VECM approach provides a meaningful structural relationship between sales and some economic indicators; however, one problem in this analysis is the limited number of economic indicators tested. The suggestion for future work with this analysis is to test the relationship of automobile sales with more economic indicators. Another issue in using VECM modeling is the linearity assumption. Additional parametric and non-parametric modeling approaches can also be used to explore nonlinear and nonstationary relationships among automobile sales and related variables.

CHAPTER VI

LONG-TERM AUTOMOBILE DEMAND PREDICTION USING INTRINSIC TIME-SCALE DECOMPOSITION

Abstract: Long-term demand prediction is crucial for industries with long product development cycles. For automobile products, where typical concept-to-release times are long (12-60 months), the process of demand evolution follows nonlinear, nonstationary dynamics, hindering accurate prediction of future automobile demand. This chapter presents the implementation details and results of using a new prediction approach based on intrinsic time-scale decomposition (ITD) to forecast automobile demand over extended time-horizons. The ITD algorithm decomposes a time series signal into a sum of components called proper rotation components and a monotonic trend. A key advantage of this approach is that this method can be used to identify long-run equilibrium relationships between automobile demand and the ITD components of economic indicators. Once the relationship is identified, this approach takes advantage of the cointegration property of the VECM model to forecast automobile demand. Hence, the long-term impact of economic indicator components on automobile demand can be quantified. The empirical results suggest that this ITD-based prediction approach can significantly improve prediction accuracy in terms of RMSE (23%) and MAPE (26%) for long-term prediction of automobile demand (12-month-ahead prediction), compared to classical and advanced time series techniques.

6.1 Introduction

Demand prediction is an essential part of any business activity. For industries with long product development cycles, long-term demand prediction serves as an input to many business decisions that affect profitability. As in the case of automobile products, where typical concept-to-release times are long (12-60 months), reliable long-term demand prediction makes an important contribution to successful service, revenue, production and inventory planning.

Automobile demand prediction has received significant attention in the literature. Many theoretical models for automobile demand prediction have been proposed [141-145]. Most of them are econometric approaches imposing the structure of a certain economic theory on the data. Only recently have a few efforts been made to address the automobile demand prediction problem using time series and data-driven approaches [135, 146, 147]. However, none of these efforts have addressed automobile demand prediction over a long-term horizon. In the economic area, some recent developments in time series techniques have been specifically designed to quantify relationships among endogenous and exogenous variables. These techniques include vector autoregressive (VAR) and vector error correction models (VECM) [75, 148].

Especially in the case of nonstationary variables, VECM has been broadly recognized as a powerful theory-driven model that can be used to describe the long-run dynamic behavior of multivariate time series. However, the choice of endogenous and exogenous variables in VECM is problematic for automobile demand prediction.

Demand is known to be influenced by many exogenous factors, such as advertising, sales promotions, retail price, and technological sophistication [149]. In the automobile market, advertising and sales promotions tend to have substantial effects; however, the effects on demand are rarely persistent [150]. To select variables for automobile demand prediction, some economic indicators, such as the Consumer Price Index (CPI) and the unemployment rate, have been

suggested as having persistent effects on automobile demand, but the results of empirical models show that long-run equilibrium relationships among these economic indicators and automobile demand do not exist. Considering the nonlinear behavior of these economic indicators, one would expect that a statistical hypothesis test of long-run equilibrium using linear assumption may perform poorly due to model misspecification.

This study presents a new prediction approach based on the intrinsic time-scale decomposition (ITD) technique, a recently developed nonparametric decomposition technique for signals that are nonlinear and/or nonstationary in nature. ITD decomposes a signal into a sum of components called proper rotation components and a monotonic trend. One advantage of the ITD method is that it requires very limited assumptions about the data. This advantage makes ITD suitable for nonlinear data from unknown underlying processes. ITD also provides unbiased decomposing components, unlike parametric algorithms which require parameter estimation. The key aspect of this new prediction approach is to develop a methodology to identify a causal and long-run equilibrium relationship from variables of interest and ITD components of related indicators and to use this identified relationship for prediction. This new prediction approach can be applied to any system of nonlinear and/or nonstationary variables. In this study, the property of ITD leads to applying it to determine the long-run equilibrium relationships between automobile demand and ITD components of nonlinear economic indicators. This investigation indicates that the long-run equilibrium of automobile demand and ITD components of unemployment rates does exist, and the ITD-based method significantly improves prediction accuracy in terms of RMSE (23%) and MAPE (26%) for long-term prediction of automobile demand (12-month ahead prediction), compared to classical and advanced time series techniques. The organization of the remainder of this chapter is as follows: implementation details and empirical results are given in Section 6.2. Conclusions and suggested future work are presented in the last section of this chapter.

6.2 Implementation Details and Empirical Results

In this study, the number of monthly retail sales (motor vehicle and parts dealers) in the U.S. during the period of January 1992- December 2010 was used to represent aggregate automobile demand. The unemployment rate during the same period of time was selected as an economic indicator to help predict demand. In order to predict long-term demand, the ITD components of the unemployment rate were hypothesized to have a long-run equilibrium relationship with automobile demand. This study investigates and test hypotheses of a long-run relationship between automobile demand and the ITD components of the unemployment rate. Table 6.1 provides the details of demand and unemployment rate. For simplicity, logarithm transformation are used with both variables in subsequent analysis. The time series in the original scale of both variables are shown in Figure 6.1.

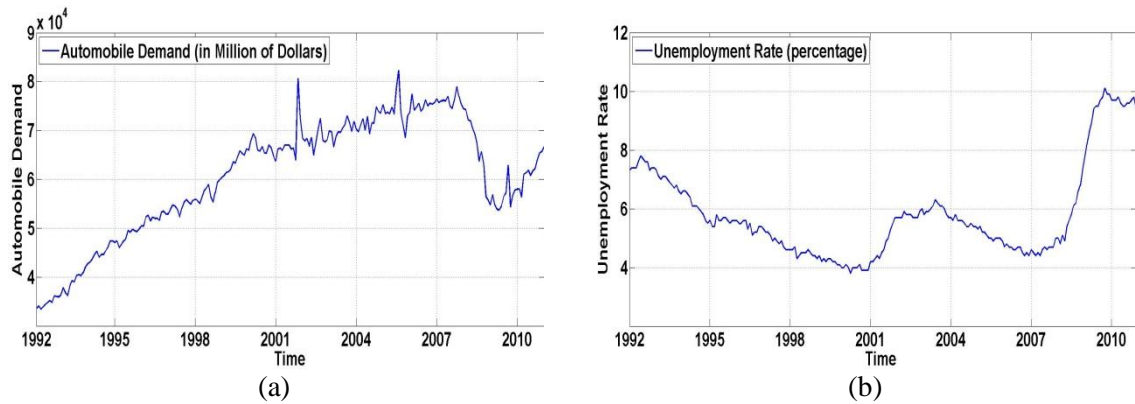


Figure 6.1: (a) Automobile Demand and (b) Unemployment rate

From the literature [151], many studies have indicated the existence of nonlinear behavior in the U.S. unemployment rate. The results of a cointegration test between unemployment rate and automobile demand, shown in Table 6.2, reveal the effect of nonlinear behavior in the unemployment rate on the long-run equilibrium relationship with automobile demand.

Table 6.1: Summary of variables

Variables	Source	Publishing Interval	Description
Automobile Demand	BLS*	Monthly	A monthly retail sales of motor vehicles and part accessories
Unemployment rate	BLS*	Monthly	A national unemployment rate (16 years or over)

* Bureau of Labor Statistics (Seasonally adjusted data)

Table 6.2: Cointegration Rank Tests (Automobile Demand and Unemployment Rate)

H_0 : Rank = r	H_1 : Rank > r	Trace Test			Maximum Eigenvalue Test		
		Trace Statistic	5%	10%	Max Statistic	5%	10%
0	0	8.6416	15.34	13.31	7.8798	14.07	12.07

Table 6.2 reports the test statistics and the corresponding asymptotic critical values of both trace and maximum Eigenvalue tests at the 5% and 10% significance levels. The empirical results of the cointegration tests show that the null hypothesis of no cointegrating vector cannot be rejected for both tests at 5% and 10% significance levels. The test results indicate that there is statistically no long-run equilibrium relationship between the unemployment rate and automobile demand. Considering the cointegrating vector requirement, VECM cannot be directly applied to these variables. As discussed in the introduction section, the cointegration tests using a linear assumption may perform poorly in the case of nonlinear variables. The next step is to apply the ITD-based prediction methodology to investigate and test the long-run equilibrium relationship between automobile demand and ITD components of the unemployment rate. The implementation details are as follows:

6.2.1 Intrinsic Time-Scale Decomposition Step

The first step in the prediction methodology is to decompose the unemployment rate using the ITD method. The ITD algorithm decomposes the unemployment rate into four proper rotation

components and a monotonic trend, as shown in Figure 6.2. The procedure using an ITD algorithm allows a perfect reconstruction of the original unemployment rate.

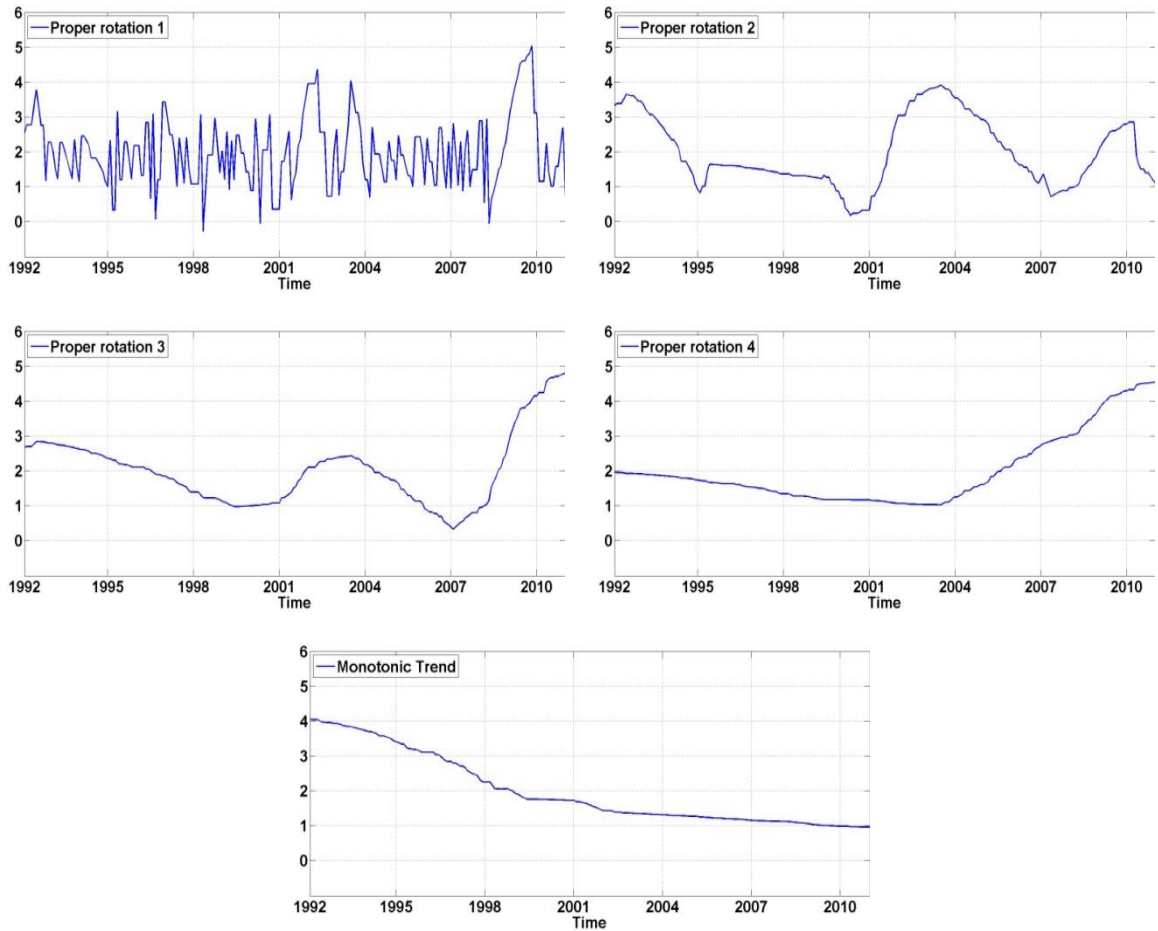


Figure 6.2: ITD components of Unemployment Rate

6.2.2 Variable Selection Step

In this second step, endogenous and exogenous variables for VECM were selected from automobile demand and ITD components of the unemployment rate obtained from the first step. Stationarity characteristics of automobile demand and each ITD component of the unemployment rate are identified using unit root tests. Exogenous variables were selected based on the results from weak exogeneity tests. In order to conduct the augmented Dickey-Fuller unit root test, it is important to use the correct number of lags in the equation. The appropriate lag length is selected

based on a general-to-specific methodology. The procedure begins with estimating an autoregressive model with a relatively long lag length. The idea is to pare down the model by the usual t -tests and/or F -tests. If the t -statistic on the last lag (p^*) is insignificant at some specified significance level, reestimate the model with a lag length of p^*-1 , then repeat the process until the last lag is significantly different from zero. The ADF test results with optimal lag length selected using a general-to-specific methodology are shown in Table 6.3.

From the ADF unit root test results, only proper rotation 2 (H2) is statistically stationary the 1% significance level. The null hypothesis of a unit root cannot be rejected for automobile demand and all other ITD components of the unemployment rate at the 1% significance level. Applying first difference to all variables except H2, the ADF test statistics reject the null hypothesis of a unit root for all differenced variables at the 5% significance level. Based on the results of the ADF tests, all variables, excluding H2, are treated as nonstationary I(1) series. Exogenous variables for the VECM are based on the results of weak exogeneity tests on I(1) series as shown in Table 6.4.

Table 6.3: ADF Unit Root Test of Original and First Differenced Variables

Variables	Optimal Lag Length	Unit Root with Drift		Unit Root with Drift and Deterministic Time Trend	
Original Variables	p^*	τ_μ	$\Pr < \tau_\mu$	τ_τ	$\Pr < \tau_\tau$
Automobile Demand (AD)	3	-3.21	0.0213	-1.70	0.75
Proper rotation 1 (H1)	4	-4.83	0.0001	-4.90	0.0004
Proper rotation 2 (H2)	4	-2.71	0.0751	-2.69	0.2405
Proper rotation 3 (H3)	5	-2.51	0.1146	-2.33	0.4130
Proper rotation 4 (H4)	5	0.14	0.9683	-0.66	0.9739
Monotonic Trend (R)	5	-2.62	0.0905	-0.68	0.9725
First Differenced Variables	p^*	τ_μ	$\Pr < \tau_\mu$	τ_τ	$\Pr < \tau_\tau$
ΔAD	6	-4.74	0.0002	-5.19	0.0002
$\Delta H2$	3	-4.34	0.0006	-4.34	0.0033
$\Delta H3$	3	-2.99	0.0377	-3.45	0.0475
$\Delta H4$	3	-2.97	0.0395	-4.36	0.0030
ΔR	4	-3.20	0.0216	-4.14	0.0065

Table 6.4: Weak Exogeneity Tests (Selected Variables)

Variables	Degree of Freedom	$\chi^2(4)$	$\Pr > \chi^2$
AD	4	22.74	0.0001
H2	4	24.25	<.0001
H3	4	12.43	0.0144
H4	4	23.95	<0.0001
R	4	6.58	0.1595

The weak exogeneity test begins with the vectors of the five variables (AD, H2, H3, H4 and R) in Table 6.3. For a system of five variables, there can exist at most four cointegrating vectors. The results in Table 6.4 show that the null hypothesis of weak exogeneity cannot be rejected for the monotonic trend (R).

Table 6.5: Weak Exogeneity Tests (Re-estimate)

Variables	Degree of Freedom	$\chi^2(4)$	$\Pr > \chi^2$
AD	3	16.95	0.0007
H2	3	12.01	0.0074
H3	3	8.53	0.0363
H4	3	19.82	0.0002

Green-slade [89] suggest re-estimating the model, using only rejected weakly exogenous variables as endogenous variables, to avoid sensitivity in the model specification. Table 6.5 shows the results of re-estimating the weak exogeneity tests of four endogenous variables, excluding the monotonic trend. The test statistics continue to reject the null hypothesis of weak exogeneity at the 5% significance level for all four variables. Based on the results of unit root and weak exogeneity tests, automobile demand (AD), and three ITD components of the unemployment rate (H2, H3, and H4) were selected as endogenous variables. For exogenous variables, due to a possible collinearity issue between automobile demand (AD) and the monotonic trend (R), only proper rotation 1 (H1) was selected as an exogenous variable.

The existence of a long-run equilibrium relationship among endogenous variables was investigated; Table 6.6 reports the test statistics and corresponding asymptotic critical values at 5% and 10% significance levels for the cointegration rank tests. The test results show that the null hypothesis of one cointegrating vector cannot be rejected at both 5% and 10% significance levels. Statistically, there is a long-run equilibrium relationship or cointegrating vector among four endogenous variables (AD, H2, H3 and H4) as shown in Eq. (6.1).

$$\beta \cdot AD = \beta_{H2} \cdot H2 + \beta_{H3} \cdot H3 + \beta_{H4} \cdot H4 \quad (6.1)$$

The underlying processes of these variables are random in the short term but tend to move together over a long-term horizon. Since all four variables are $I(1)$ series, the cointegrating vector in Eq. (6.19) is a stationary process ($I(d-b) = I(0)$ where $d = b = 1$). Long-run parameter (β) estimates of the cointegrating vector are presented in Table 6.7, where the parameter estimate for automobile demand is normalized ($\beta=1$).

Table 6.6: Cointegration Rank Tests (Automobile Demand and Selected ITD Components of the Unemployment Rate)

H ₀ : Rank = r	H ₁ : Rank > r	Trace Test			Maximum Eigenvalue Test		
		Trace Statistic	5%	10%	Max Statistic	5%	10%
0	0	52.5057	47.21	43.84	29.5151	27.07	24.73
1	1	22.9906	29.38	26.70	14.4999	20.97	18.60

Table 6.7: Long-Run Parameter Beta Estimates of Cointegrating Vector (AD, H2, H3 and H4)

Variable	AD	H2	H3	H4
Beta	β	B_{H2}	B_{H3}	B_{H4}
Parameter Estimate	1	0.41389	-0.46206	0.68588

6.2.3 Prediction Step

The results of steps 1 and 2 show that VECM is applicable to automobile demand and the ITD components of the unemployment rate since the cointegrating vector of these variables can be

identified. The ITD-based prediction using VECMX with 4 endogenous (AD, H2, H3 and H4) and 1 exogenous (H1) variable was estimated. To evaluate the forecasting performance of the ITD-based prediction using VECMX, the out-of-sample forecasting of this model is compared with those from three rival models. The first model is an autoregressive integrated moving average (ARIMA) model, the most general class of univariate time series models which can be applied to nonstationary variables. The second model is an autoregressive integrated moving average with exogenous variables (ARIMAX) model. The ARIMAX model of automobile demand is selected as a baseline model since it allows the use of the unemployment rate as the exogenous variable; however, it cannot handle feedback from one variable to the other. The third model is the vector autoregressive (VAR) model. The VAR model can handle feedback from one variable to the other, but does not impose cointegrating restrictions as in the VECMX. It treats automobile demand and the unemployment rate symmetrically. Two model selection criteria were selected for out-of-sample comparison: root mean square error (RMSE) and mean absolute percentage error (MAPE). The out-of-sample forecasting initially identified each model specification over the sample period of January 1992 to May 2007, then each model was used to generate forecasts of four-, eight- and twelve-step ahead predictions. The sample was then rolled forward for one month, and another set of four- to twelve-step ahead predictions was generated. As a result, 24 four-, eight- and twelve-step ahead predictions were obtained. Table 6.8 provides RMSE and MAPE values of four-, eight- and twelve-step ahead predictions for automobile demand (AD) for each model. Figures 6.4(a) and 6.4(b) show the out-of-sample forecasting comparison of the four models using RMSE and MAPE.

As seen in Table 6.8, the ITD-based prediction model is the best model in terms of RMSE and MAPE. It significantly improves the accuracy of all four-, eight-, and twelve-step ahead automobile demand predictions. For four-step ahead prediction, the ITD-based prediction model reduces RMSE and MAPE by 25% and 30% respectively, compared to the second best model

(VAR model). For 8-step ahead prediction, the ITD-based prediction model reduces RMSE and MAPE by 24% and 25% respectively, compared to VAR model, and for 12-step ahead prediction, ITD reduces RMSE and MAPE by 23% and 26% respectively, compared to the VAR model.

Table 6.8: Out-of-Sample Forecasting Comparison

Model	4-step ahead prediction		8-step ahead prediction		12-step ahead prediction	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARIMA	0.0846	0.0071	0.1562	0.0132	0.2182	0.0190
ARIMAX	0.0859	0.0072	0.1585	0.0134	0.2206	0.0192
VAR	0.0772	0.0064	0.1440	0.0121	0.1999	0.0173
ITD-based Prediction	0.0578	0.0045	0.1100	0.0091	0.1531	0.0128

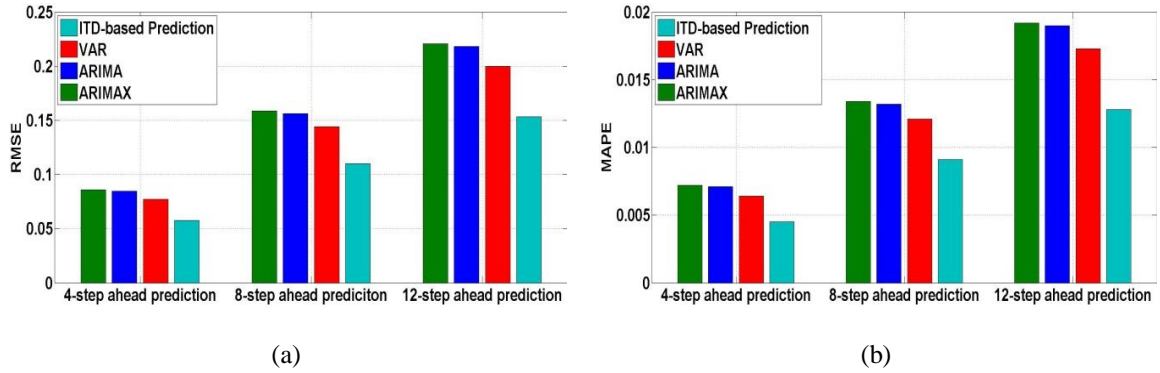


Figure 6.3: (a) Out-of-Sample Forecasting Comparison using RMSE and (b) Out-of-Sample Forecasting Comparison using MAPE

6.3 Conclusions and Suggested Future Work

In this study, a new prediction approach based on intrinsic time-scale decomposition (ITD) was proposed for a long-term prediction application using nonlinear and/or nonstationary variables. ITD is a recently developed nonparametric algorithm aiming to enhance the analysis of nonlinear and nonstationary signals. The ITD method represents a nonlinear and/or nonstationary signal as

a summation of components called proper rotation and a monotonic trend. The advantage of the ITD-based prediction approach is that, from a system of variables with no long-run equilibrium relationships or cointegrating vectors, this approach can identify the cointegrating vector of one variable and the ITD components of other variables. In this study, the ITD-based prediction approach was applied to a long-term prediction of automobile demand with the aid of an economic indicator. Considering that the economic indicator is a combination of multiple components with different characteristics resulting from factors underlying each component, this property of ITD has led to a prediction application using the long-run equilibrium relationship or cointegrating vector of automobile demand and the ITD components of an economic indicator. The empirical results show that automobile demand has a long-run equilibrium relationship with some ITD components of the unemployment rate and that the ITD-based prediction approach can significantly improve the accuracy of long-term prediction of automobile demand.

To further improve prediction accuracy using the ITD algorithm, one may consider the edge effect of ITD discussed in Section 3.3 as suggested future work. The edge effect of ITD on the most recent data is confined to the interval of the last two extremas because the most recent data point of the signal is treated as an extremum, which may not be true when more data become available. In a prediction application, the edge effect of ITD has a significant effect on prediction accuracy. One possible alternative for solving this issue is to construct the decomposition beyond the last available data point. However, this extension of the data is risky because the edge effect will depend on the accuracy of the extended data. Currently, no simple approach or general theory exists to solve the edge effect issue of ITD.

With the property that a proper rotation between two consecutive local extrema is monotonic, ITD has an advantage in the analysis, which is that the whole data span need not be extended. The only information needed are the value and location of the next two extrema. Even with this

advantage of ITD, the task of improving prediction accuracy by solving the edge effect issue remains challenging.

CHAPTER VII

COMPARISON OF NONLINEAR AND NONSTATIONARY FORECASTING MODELS

This chapter presents a summary of the comparison of different forecasting models, including ARCH/CARCH, MRTAR (Multi-regime threshold AR), WNN (wavelet neural network), NNM (nearest neighborhood model), PF, ITD, LGP (local recurrence-based GP) and DPMG, in terms of model capacity, such as nonlinear form (e.g., linear trend), nonstationary form (e.g., changed variance), computational (training) complexity, and noise effect. Also, two case studies using these forecasting models are presented.

7.1 Comparison of Nonlinear and Nonstationary Forecasting Models

Table 7.1 presents a summary of the performance of different nonlinear nonstationary prediction models reported in prior empirical studies. Training complexity (tc) represents the computational complexity. It is a function of various factors, including model order, number of inputs, input dimension and so on. Here, n is the number of inputs, p and q are the model orders, n_c is the number of regimes or clusters, n_n is the number of neurons in NN, and n_x is the state dimension. Based on the investigation, the functional decomposition models can accommodate large noise variances in the underlying system, and MRTAT and PF models can incorporate exogenous variables as the unobserved state. Compared to classical GP models, LGP and DPMG models can be applied to nonstationary time series and reduce the computational overhead, as the number of training points in each segment or cluster is reduced.

Table 7.1: Summary of the performance of different prediction models in prior empirical studies [152]

Prediction models	Nonlinearity	Nonstationarity	Training complexity	Remarks
ARCH/GARCH	Linear	Time-varying volatility	$tc(n, p, q)$	Linear structures; can model varying second moments; sensitive to noise
MRTAR	Piecewise linear	Switch according to time and indicator	$tc(n_c, p, q, n)$	Exogenous variables used to indicate switch
Wavelet-NN	Between lagged variables	Nonstationarity dispersed over scales	$tc(n_n, n)$	Can separate relatively large noise
PF	Between lagged variables	Adaptive parameter update	$tc(n_x, n)$	Posterior approximation; MCMC computations are cumbersome
NNM	Local linearity/nonlinearity	Local stationarity	$tc(n)$	Sensitive to neighborhood threshold; can accommodate large noise level
ITD	Decomposed into monotonic trend	Decomposed into rotation components	$tc(n)$	Construct piecewise linear baseline signal; reduce boundary effect in EMD; reduced sensitivity to noise
LGP	Between lagged variables	Piecewise stationary	$tc(n)$	Recurrence based; sensitive to threshold; GP requires relatively low noise level
DPMG	Between lagged variables	Piecewise stationary	$tc(n_c, n)$	Dirichlet process clustering; GP accuracy sensitive to noise

7.2 Case Study I: Real-time Throughput Forecasting in an Automotive Assembly Line

In this case study, a time series representing the number of parts produced (also called the throughput) in an 8-hour shift from a station of an automotive assembly line [44] was used. A comparison of forecasting accuracies of different models is summarized in Table 7.2.

Table 7.2: Comparison of forecasting and accuracy of automotive assembly line throughput [152]

	ARMA	LGP	LRM	LMM	EMD	DPMG
Forecasting horizon	2	3	3	5	6	5
R^2 (first step)	0.10	0.61	0.53	0.60	0.60	0.59

Here, the PF/RPNN model did not converge because of high nonstationarity and data sparsity. All methods tested except ARMA provide comparable 1-step forecasting accuracy (R^2 in the range of 0.5-0.6), as they are designed to capture the nonlinear and nonstationary evolution. The linear

structure of ARMA model is evidently inadequate to capture the complex throughput variation. Among the nonlinear and nonstationary forecasting methods, EMD provides the highest long term (6-step) forecasting accuracy as some of the IMFs exhibit a near-stationary evolution. LGP has the best prediction accuracy in terms of the second moment, and over 85% of the realized values are within 2-sigma (estimated) limits of the predictions likely due to effective partitioning of the state space into local near-stationary segments.

7.3 Case Study II: End Point Detection in Chemical Mechanical Planarization Processes

The semiconductor industry relies on chemical mechanical planarization (CMP) processes to polish wafer surfaces to meet the strongest surface roughness, flatness, and defect-control regularizations. Various sensors can be used during the polishing process for *in situ* monitoring and control of material removal rate (MRR) and surface quality [35, 153]. The sensor signals were gathered from the process under various conditions, and the dynamics underlying the measured signals were found to be nonlinear and nonstationary. Fourteen different features capturing the various complex patterns of the signals were extracted, and 9 principal components of these features were used to predict MRR, necessary for timely control of wafer height. An NN model was used to predict MRR from the forecasts of the first 9 principal features. An evolution of the first principal feature suggests significant nonstationarity and aperiodicity of the dynamics. Table 7.3 summarizes the accuracy of forecasting first principal feature and MRR using alternative methods. While PF can adaptively adjust to the nonlinear dynamics and provide the highest accuracy for forecasting the feature evolution as well as MRR, EMD offers higher accuracy for relatively long-term forecasting, as the decomposed IMFs represent the long-term trend in the time series. Also, 91% of the realized feature values were within 2-sigma (estimated) limits of DPMG forecasts, thus facilitating accurate endpoint detection to mitigate over- and under-polishing defects in semiconductor wafers.

Table 7.3: Comparison of forecasting horizon and accuracies for the first principal feature and
MRR [152]

	ARMA	LGP	LRM	PF	RPNN	EMD	DPMG
Forecasting horizon	1	3	2	3	1	4	3
R^2 (first step)	0.07	0.53	0.56	0.80	0.10	0.75	0.61
MRR estimation R^2 (first step)	0.01	0.63	0.67	0.89	0.05	0.84	0.71

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

Recent advancements in sensor, computing, and communication technologies and the consequent availability of abundant data sources in the form of time series can transform the way real-world complex systems are monitored and controlled. Time series forecasting of the evolution of complex systems is considered one of the emerging challenges of modern science. Effective prediction of future states of a complex system from time series remains a challenge, mainly because of diverse combinations of the nonlinear and nonstationary dynamic behaviors exhibited by these systems. This study presents a novel approach to nonlinear and nonstationary time series forecasting. Research methodologies were developed using nonlinear decomposition techniques and a battery of advanced statistical methods. The major conclusions and recommendations for future work follow.

8.1 Conclusions

In this study, two new nonlinear and nonstationary prediction methodologies based on nonlinear decomposition techniques were presented: empirical mode decomposition (EMD)-based prediction methodology and intrinsic time-scale decomposition (ITD)-based prediction methodology. EMD-based prediction addresses the challenge (edge artifacts limits) of using EMD for prediction applications. This methodology was developed specifically for univariate nonlinear and nonstationary time series forecasting. A two-step algorithm has been

designed to address the edge artifacts limits, resulting in a significant improvement in prediction accuracy as shown in the case of a customer willingness-to-pay prediction application.

ITD-based prediction methodology was developed specifically for multivariate time series prediction. It utilizes the structural relationship identification (SRI) methodology to identify relationship of variables in the system prior to adapting the ITD technique for prediction. SRI methodology addresses the identification problem of VAR and VECM types of models. It captures the interdependencies of a multivariate time series system and provides evidence of the dynamic behavior and structural relationships among variables. It also allows a feedback relationship, which can be used to capture more complex interactions among variables.

The ITD technique overcomes the limitations of the EMD approach for nonlinear and nonstationary time series modeling. It decomposes time series into proper rotation components with defined frequency and amplitude and a monotonic trend. One advantage of using ITD over EMD for prediction applications is that the edge effect of the ITD process is confined to the interval $[0, \tau_2]$ at the beginning and $[\tau_k, \tau_{k+2}]$ at the ending of each proper rotation and monotonic trend. The results of using ITD-based prediction methodology for an automobile demand prediction application show that the ITD-based prediction approach can significantly improve the accuracy of long-term predictions of automobile demand.

8.2 Future Work

In the area of nonlinear and nonstationary time series forecasting, future work will be to continue improving the functional decomposition model for nonlinear and nonstationary time series. The continued improvements and prospective studies will be as follows:

For the ITD-based prediction methodology, SRI methodology has been proved to identify long-run equilibrium interrelationships among variables in the system. However, there is some disadvantage in implementing this methodology in practice. A well-known problem of VAR and

VECM is the over-parameterization issue, which is a prohibitively large number of parameters to be estimated. One way to address this problem is to impose theory-based weak exogeneity assumptions on variables (as used in ITD-based prediction methodology). The number of equations in the model can be reduced if variables are treated as weakly exogenous in the model. However, imposing the test alone may not be sufficient in cases of large datasets.

Recently, in the field of data mining, significant efforts have been made to address the issue of an excessive number of correlated factors. Many dimensional and variable selection techniques have been proposed to solve the problem. In the case of dimensional reduction techniques, although they are useful to retain a significant portion of explained variance with a reduced number of factors, interpretations of the results are no longer straightforward because the components from dimensional reduction techniques are a combination of all of the selected variables. The results are not easily explained and may not be applicable in the context of econometrics. Considering the disadvantage of dimensional reduction techniques, variable selection methods have gained significantly more attention for analysts in the sense that the results from variable selection methods can be used and explained directly. To improve the prediction performance of the ITD-based prediction, a future study may be to use data mining techniques to solve the over-parameterization issue in SRI methodology.

In addition to variable selection techniques, one may consider the edge effect of ITD as discussed in Section 3.3 for future research. The edge effect of ITD on the most recent data is confined to the interval of the last two extrema because the most recent data point of the signal is treated as an extremum, which may not be true when more data become available. In a prediction application, the edge effect of ITD has a significant effect on prediction accuracy. One possible alternative to solve this issue is to construct the decomposition beyond the last available data point. However, this extension of the data is risky because the edge effect will depend on the

accuracy of the extended data. Currently, no simple approach or general theory exists to solve the edge effect issue of ITD.

With the property that a proper rotation between two consecutive local extrema is monotonic, ITD has an advantage in the analysis, which is that the whole data span need not be extended. The only information needed are the value and location of the next two extrema. Even with this advantage of ITD, the task of improving prediction accuracy by solving the edge effect issue remains challenging.

REFERENCES

- [1] C. Kumar, *et al.*, "21st-Century Physics: Grand Challenges," *Journal of the Federation of American Scientists*, vol. 56, pp. 9-11, 2003.
- [2] M. G. Frei and I. Osorio, "Intrinsic time-scale decomposition: Time–frequency–energy analysis and real-time filtering of non-stationary signals," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, vol. 463, pp. 321-342, 2007.
- [3] N. E. Huang, *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903-995, March 8, 1998.
- [4] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, pp. 443-473, 2006.
- [5] F. Battaglia and M. K. Protopapas, "Time-varying multi-regime models fitting by genetic algorithms," *Journal of Time Series Analysis*, vol. 32, pp. 237-252, 2011.
- [6] J. Fan and Q. Yao, *Nonlinear Time Series*, vol. 2: New York: Springer, 2002.
- [7] H. Stark and J. W. Woods, *Probability, Statistics, and Random Processes for Engineers*, Upper Saddle River, NJ : Pearson, 2012.
- [8] D. Koutsoyiannis and A. Montanari, "Statistical analysis of hydroclimatic time series: Uncertainty and insights," *Water Resources Research*, vol. 43, 2007.
- [9] M. G. Genton and K. R. Thompson, "Skew-elliptical time series with application to flooding risk," in *Time Series Analysis and Applications to Geophysical Systems*, New York: Springer, pp. 169-185, 2004.

- [10] H. Tong, *Non-linear Time Series: A Dynamical System Approach*, Oxford, UK: Oxford University Press, 1990.
- [11] P. K. Yadav, *et al.*, "Threshold autoregressive modeling in finance: The price differences of equivalent assets1," *Mathematical Finance*, vol. 4, pp. 205-221, 1994.
- [12] G. C. Tiao and R. S. Tsay, "Some advances in non-linear and adaptive modelling in time-series," *Journal of Forecasting*, vol. 13, pp. 109-131, 1994.
- [13] T. Teräsvirta, "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, vol. 89, pp. 208-218, 1994.
- [14] A. Sa-ngasoongsong, *et al.*, "Multi-step sales forecasting in automotive industry based on structural relationship identification," *International Journal of Production Economics*, vol. 140, pp. 875-887, 2012.
- [15] D. Lowe and A. Webb, "Time series prediction by adaptive networks: A dynamical systems perspective," *IEEE Proceedings in Radar and Signal Processing*, pp. 17-24, 1991.
- [16] H. S. Hippert, *et al.*, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, pp. 44-55, 2001.
- [17] N. Kohzadi, *et al.*, "A comparison of artificial neural network and time series models for forecasting commodity prices," *Neurocomputing*, vol. 10, pp. 169-181, 1996.
- [18] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, pp. 246-257, 1991.
- [19] G. P. Zhang, "Neural networks for time-series forecasting," in *Handbook of Natural Computing*, Springer Berlin Heidelberg, 2012, pp. 461-477.

- [20] A. Laepes and R. Farben, "Nonlinear signal processing using neural networks: Prediction and system modelling," *Technical report*, Los Alamos National Laboratory, Los Alamos, NM, 1987.
- [21] C. De Groot and D. Wuertz, "Analysis of univariate time series with connectionist nets: A case study of two classical examples," *Neurocomputing*, vol. 3, pp. 177-192, 1991.
- [22] C. M. Kuan and T. Liu, "Forecasting exchange rates using feedforward and recurrent neural networks," *Journal of Applied Econometrics*, vol. 10, pp. 347-364, 1995.
- [23] G. Grudnitski and L. Osburn, "Forecasting S&P and gold futures prices: An application of neural networks," *Journal of Futures Markets*, vol. 13, pp. 631-643, 1993.
- [24] K. P. Rao, *et al.*, "Real-time identification of incipient surface morphology variations in ultra-precision machining process," *ASME Journal of Manufacturing Science and Engineering*, vol. 136(2): 021008, pp. 1-11, 2013.
- [25] J. M. P. Menezes Jr and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing*, vol. 71, pp. 3335-3343, 2008.
- [26] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199-222, 2004.
- [27] R. M. Balabin and E. I. Lomakina, "Support vector machine regression (SVR/LS-SVM)—An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data," *Analyst*, vol. 136, pp. 1703-1712, 2011.
- [28] L. Cao, "Support vector machines experts for time series forecasting," *Neurocomputing*, vol. 51, pp. 321-339, 2003.
- [29] S. Mukherjee, *et al.*, "Nonlinear prediction of chaotic time series using support vector machines," in *Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing*, pp. 511-520, 1997.

- [30] K. W. Lau and Q. H. Wu, "Local prediction of non-linear time series using support vector regression," *Pattern Recognition*, vol. 41, pp. 1539-1547, 2008.
- [31] T. V. Gestel, *et al.*, "Financial time series prediction using least squares support vector machines within the evidence framework," *IEEE Transactions on Neural Networks*, vol. 12(4), pp. 809-821, 2001.
- [32] L. Cao and Q. Gu, "Dynamic support vector machines for non-stationary time series forecasting," *Intelligent Data Analysis*, vol. 6, pp. 67-83, 2002.
- [33] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257-286, 1989.
- [34] Z. Wang, *et al.*, "An extended Kalman filtering approach to modeling nonlinear dynamic gene regulatory networks via short gene expression time series," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 6, pp. 410-419, 2009.
- [35] Z. Kong, *et al.*, "Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear Bayesian analysis and statistical modeling," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, pp. 316-327, 2010.
- [36] N. J. Gordon, *et al.*, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140(2), pp. 107-113, 1993.
- [37] D. Yankov, *et al.*, "Ensembles of nearest neighbor forecasts," in *Machine Learning: ECML 2006*, Springer Berlin Heidelberg, pp. 545-556, 2006.
- [38] R. Mehrotra and A. Sharma, "Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach," *Advances in Water Resources*, vol. 29, pp. 987-999, 2006.

- [39] N. Z. A. Hamid and M. S. M. Noorani, "Modeling of prediction system: An application of nearest neighbor approach to chaotic data," *Applied Mathematics and Computational Intelligence*, vol. 2(1), pp. 137-148, 2013.
- [40] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, England: Cambridge University Press, 1997.
- [41] B. Finkenstädt and P. Kuhbier, "Forecasting nonlinear economic time series: A simple test to accompany the nearest neighbor approach," *Empirical Economics*, vol. 20, pp. 243-263, 1995.
- [42] M. Casdagli, "Recurrence plots revisited," *Physica D: Nonlinear Phenomena*, vol. 108, pp. 12-44, 1997.
- [43] M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D: Nonlinear Phenomena*, vol. 35, pp. 335-356, 1989.
- [44] H. Yang, *et al.*, "Local recurrence based performance prediction and prognostics in the nonlinear and nonstationary systems," *Pattern Recognition*, vol. 44, pp. 1834-1840, 2011.
- [45] S. T. S. Bukkapatnam, *et al.*, "Towards Prediction of nonlinear and nonstationary evolution of customer preferences using local Markov models," in *Product Research*, N. R. S. Raghavan and J. A. Cafeo, Eds., Springer Netherlands, pp. 271-287, 2009.
- [46] J. McNames, "A nearest trajectory strategy for time series prediction," *Proceedings of the International Workshop on Advanced BlackBox Techniques for Nonlinear Modeling*, pp. 112-128, 1998.
- [47] S. Regonda, *et al.*, "Local polynomial method for ensemble forecast of time series," *Nonlinear Processes in Geophysics*, vol. 12, pp. 397-406, 2005.
- [48] S. J. Gershman and D. M. Blei, "A tutorial on Bayesian nonparametric models," *Journal of Mathematical Psychology*, vol. 56, pp. 1-12, 2012.
- [49] C. E. Rasmussen, "Gaussian processes for machine learning," *Advanced Lectures on Machine Learning*, Springer Berlin Heidelberg, pp. 63-71, 2004.

- [50] S. Brahim-Belhouari and A. Bermak, "Gaussian process for nonstationary time series prediction," *Computational Statistics & Data Analysis*, vol. 47, pp. 705-712, 2004.
- [51] C. Plagemann, *et al.*, "Nonstationary Gaussian process regression using point estimates of local smoothness," in *Machine learning and Knowledge Discovery in Databases*, Springer Berlin Heidelberg, pp. 204-219, 2008.
- [52] G. Rilling, *et al.*, "On empirical mode decomposition and its algorithms," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP*, pp. 8-11, 2003.
- [53] X. Wang, *et al.*, "Time series forecasting energy-efficient organization of wireless sensor networks," *Sensors*, vol. 7, pp. 1766-1792, 2007.
- [54] X. Xu, *et al.*, "Forecasting demand of commodities after natural disasters," *Expert Systems with Applications*, vol. 37, pp. 4313-4317, 2010.
- [55] A. Stroeer, *et al.*, "Methods for detection and characterization of signals in noisy data with the Hilbert-Huang transform," *Physical Review D*, vol. 79(12): 124022, pp. 1-10, 2009.
- [56] A. Sa-ngasoongsong and S. T. Bukkapatnam, "Willingness-to-pay prediction based on empirical mode decomposition," in *Industrial Engineering Research Conference, Reno, NV*, 2011.
- [57] X. An, *et al.*, "Short-term prediction of wind power using EMD and chaotic theory," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, pp. 1036-1042, 2012.
- [58] A. Wolf, *et al.*, "Determining Lyapunov exponents from a time series," *Physica D: Nonlinear Phenomena*, vol. 16, pp. 285-317, 1985.
- [59] C.-J. Yu, *et al.*, "Frequency spectrum prediction method based on EMD and SVR," in *Eighth International Conference on Intelligent Systems Design and Applications (ISDA'08)*, pp. 39-44, 2008.

- [60] J. P. Li, *et al.*, "Country risk forecasting for major oil exporting countries: A decomposition hybrid approach," *Computers & Industrial Engineering*, vol. 63, pp. 641-651, 2012.
- [61] Z. Zhu, *et al.*, "Hybrid of EMD and SVMs for short-term load forecasting," in *IEEE International Conference on Control and Automation (ICCA 2007)*, pp. 1044-1047, 2007.
- [62] P. Yang, *et al.*, "The prediction of non-stationary climate series based on empirical mode decomposition," *Advances in Atmospheric Sciences*, vol. 27, pp. 845-854, 2010.
- [63] T. Lee and T. Ouarda, "Prediction of climate nonstationary oscillation processes with empirical mode decomposition," *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 116, pp. 1-15, 2011.
- [64] N. E. Huang, "Introduction to the Hilbert-Huang transform and its related mathematical problems," *Interdisciplinary Mathematics*, vol. 5, pp. 1-26, 2005.
- [65] L. Hong, "Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition," *Energy Procedia*, vol. 5, pp. 1333-1340, 2011.
- [66] C. Fu, "Forecasting exchange rate with EMD-based support vector regression," in *International Conference on Management and Service Science (MASS)*, pp. 1-4, 2010.
- [67] R.J. Martis, *et al.*, "Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction " *International Journal of Neural Systems*, vol. 23(5), 1350023, pp. 1-16, 2013.
- [68] A. Sa-ngasoongsong, "Long-term demand prediction using long-run equilibrium relationship of intrinsic time-scale decomposition components," in *Industrial and Systems Engineering Research Conference (ISERC)*, Orlando, FL, 2012.
- [69] C. Gerald and P. Wheatley, *Applied Numerical Analysis*: Addison-Wesley, 1994.
- [70] N. E. Huang and S. S. P. Shen, *Hilbert-Huang Transform and Its Applications*. London: World Scientific, 2005.

- [71] K. T. Coughlin and K. K. Tung, "11-Year solar cycle in the stratosphere extracted by the empirical mode decomposition method," *Advances in Space Research*, vol. 34, pp. 323-329, 2004.
- [72] P. Flandrin, *et al.*, "Detrending and denoising with empirical mode decompositions," in *Proceedings of the European Signal Processing Conference*, Vienna, Austria, vol. 2, pp. 1581-1584, 2004.
- [73] W. Zhaohua and E. H. Norden, "A study of the characteristics of white noise using the empirical mode decomposition method," *The Royal Society*, vol. 460, pp. 1597-1611, 2004.
- [74] W. Zhaohua, *et al.*, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104(38), pp. 14889-14894, 2007.
- [75] M. G. Dekimpe, *et al.*, "Time-series models in marketing," in *Handbook of Marketing Decision Models*. vol. 121, B. Wierenga, Ed., Springer US, pp. 373-398, 2008.
- [76] M. G. Dekimpe and D. M. Hanssens, "The persistence of marketing effects on sales," *Marketing Science*, vol. 14, pp. 1-21, 1995.
- [77] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, pp. 427-431, 1979.
- [78] D. A. Dickey and W. A. Fuller, "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, vol. 49, pp. 1057-1072, 1981.
- [79] J. Y. Park, "Testing for Unit Roots and Cointegration by Variable Addition," *Advances in Econometrics*, vol. 8(2), pp. 107-133, 1990.
- [80] J. Y. Park and B. Choi, "A new approach to testing for a unit root," Working paper no. 88-23, Cornell University, Ithaca, NY, 1988.

- [81] S. Johansen, "Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models," *Econometrica*, vol. 59, pp. 1551-1580, 1991.
- [82] C. A. Sims, "Macroeconomics and reality," *Econometrica*, vol. 48, pp. 1-48, 1980.
- [83] J. Søren, "Testing weak exogeneity and the order of cointegration in UK money demand data," *Journal of Policy Modeling*, vol. 14, pp. 313-334, 1992.
- [84] R. F. Engle, *et al.*, "Exogeneity," *Econometrica*, vol. 51, pp. 277-304, 1983.
- [85] J.-P. Urbain, "On weak exogeneity in error correction models.," *Oxford Bulletin of Economics and Statistics*, vol. 54, pp. 187-207, 1992.
- [86] C. W. J. Granger, *et al.*, "A bivariate causality between stock prices and exchange rates: Evidence from recent asian flu," *The Quarterly Review of Economics and Finance*, vol. 40, pp. 337-354, 2000.
- [87] C. W. J. Granger, "Causality, cointegration, and control," *Journal of Economic Dynamics and Control*, vol. 12, pp. 551-559, 1988.
- [88] C. Hiemstra and J. D. Jones, "Testing for linear and nonlinear Granger causality in the stock price-volume relation," *The Journal of Finance*, vol. 49, pp. 1639-1664, 1994.
- [89] J. V. Greenslade, *et al.*, "On the identification of cointegrated systems in small samples: A modelling strategy with an application to UK wages and prices," *Journal of Economic Dynamics and Control*, vol. 26, pp. 1517-1537, 2002.
- [90] R. Engle and C. W. J. Granger, "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, vol. 55, pp. 251-276, 1987.
- [91] S. Johansen, "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, vol. 12, pp. 231-254, 1988.
- [92] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, pp. 599-607, 1984.
- [93] W. Enders, *Applied Econometric Time Series*, 3 ed. New York, NY: John Wiley & Sons, 2010.

- [94] S. Johansen, "A statistical analysis of cointegration for I(2) variables," *Econometric Theory*, vol. 11, pp. 25-59, 1995.
- [95] S. Johansen and K. Juselius, "Maximum likelihood estimation and inference on cointegration - with applications to the demand for money," *Oxford Bulletin of Economics and Statistics*, vol. 52, pp. 169-210, 1990.
- [96] H. Lutkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer, 2005.
- [97] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, New York, NY: John Wiley & Sons, 1980.
- [98] G. P. Dwyer, "Nonlinear time series and financial applications," Federal Reserve Bank of Atlanta, 2003.
- [99] M. G. Frei and I. Osorio, "Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals," *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences*, vol. 463, pp. 321-342, 2007.
- [100] K. Train and C. Winston, "Vehicle choice behavior and the declining market share of U.S. automakers," *International Economic Review*, vol. 48, pp. 1469-1496, 2007.
- [101] C. Breidert, "Estimation of willingness-to-pay. Theory, measurement, and application," Doctoral thesis, WU Vienna University of Economics and Business, WU Vienna University, 2005.
- [102] C. Goutam, *et al.*, "The ability of ratings and choice conjoint to predict market shares: A Monte Carlo simulation," *Journal of Business Research*, vol. 55, pp. 237-249, 2002.
- [103] L. Mohamed, *et al.*, "Modeling preference evolution in discrete choice models: A Bayesian state-space approach," *Quantitative Marketing and Economics*, vol. 4, pp. 57-81, 2006.

- [104] J. Louviere and T. Islam, "A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best-worst scaling," *Journal of Business Research*, vol. 61(9), pp. 903-911, 2008.
- [105] P. M. Garcia, "Measuring willingness-to-pay in discrete choice models with semi-parametric techniques," *Revista Latinoamericana de Desarrollo Economico*, vol. 5, pp. 83-100, 2005.
- [106] J. Z. Tabin, "Method and Apparatus for Interactive Criteria-based Commodity Comparisons," USA Patent, 2008.
- [107] Rob J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Transactions on Power Systems*, vol. 25, pp. 1142-1153, 2010.
- [108] Ruijun Dong and W. Pedrycz, "A granular time series approach to long-term forecasting and trend forecasting," *Physica A*, vol. 387, pp. 3253-3270, 2008.
- [109] C. Granger and Y. Jeon, "Long-term forecasting and evaluation," *International Journal of Forecasting*, vol. 23, pp. 539-551, 2007.
- [110] J. Freedman, *The U.S. Auto Industry: American Carmakers and the Economic Crisis*. New York, NY: The Rosen Publishing Group, 2011.
- [111] J. G. Monks, *Operations Management: Theory and Problems*, 2 ed. New York, NY: McGraw-Hill, 1987.
- [112] C.-W. Chu and G. P. Zhang, "A comparative study of linear and nonlinear models for aggregate retail sales forecasting," *International Journal of Production Economics*, vol. 86, pp. 217-231, 2003.
- [113] J. T. Luxhøj, *et al.*, "A hybrid econometric—neural network modeling approach for sales forecasting," *International Journal of Production Economics*, vol. 43, pp. 175-192, 1996.
- [114] P. Danese and M. Kalchschmidt, "The role of the forecasting process in improving forecast accuracy and operational performance," *International Journal of Production Economics*, vol. 131, pp. 204-214, 2011.

- [115] J. T. Mentzer and M. A. Moon, *Sales Forecasting Management: A Demand Management Approach*, 2 ed. Thousand Oaks, CA: Sage Publications, 2005.
- [116] J. Berkovec, "Forecasting automobile demand using disaggregate choice models," *Transportation Research Part B: Methodological*, vol. 19, pp. 315-329, 1985.
- [117] A. Greenspan and D. Cohen, "Motor vehicle stocks, scrappage, and sales," *The Review of Economics and Statistics*, vol. 81, pp. 369-383, 1999.
- [118] F. L. Mannering and K. Train, "Recent directions in automobile demand modeling," *Transportation Research Part B: Methodological*, vol. 19, pp. 265-274, 1985.
- [119] C. A. Lave and K. Train, "A disaggregate model of auto-type choice," *Transportation Research Part A: General*, vol. 13, pp. 1-9, 1979.
- [120] P. H. Franses, "Modeling new product sales: An applicaiton of cointegration analysis," *International Journal of Research in Marketing*, vol. 11, pp. 491-502, 1994.
- [121] J. S. Armstrong, *et al.*, "Sales forecasts for existing consumer products and services: Do purchase intentions contribute to accuracy?" *International Journal of Forecasting*, vol. 16, pp. 383-397, 2000.
- [122] M. Hulsmann, *et al.*, "General sales forecast models for automobile markets and their analysis," *Transactions on Machine Learning and Data Mining*, vol. 5, pp. 65-86, 2012.
- [123] L. Kunhui, *et al.*, "Time series prediction based on linear regression and SVR," in *Third International Conference on Natural Computation (ICNC 2007)*, Haikou, China, pp. 688-691, 2007.
- [124] P. H. Franses, "Fitting a Gompertz curve," *The Journal of the Operational Research Society*, vol. 45, pp. 109-113, 1994.
- [125] N. Meade, "The use of growth curves in forecasting market development—A review and appraisal," *Journal of Forecasting*, vol. 3, pp. 429-451, 1984.
- [126] J. R. Landwehr, *et al.*, "Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts," *Marketing Science*, vol. 30, pp. 416-429, 2011.

- [127] M. G. Dekimpe, *et al.*, "Long-run effects of price promotions in scanner markets," *Journal of Econometrics*, vol. 89, pp. 269-291, 1998.
- [128] K. Pauwels, *et al.*, "New products, sales promotions, and firm value: The case of the automobile industry," *The Journal of Marketing*, vol. 68, pp. 142-156, 2004.
- [129] V. R. Nijs, *et al.*, "The category-demand effects of price promotions," *Marketing Science*, vol. 20, pp. 1-22, 2001.
- [130] K. Pauwels, *et al.*, "The long-term effects of price promotions on category incidence, brand choice, and purchase quantity," *Journal of Marketing Research*, vol. 39, pp. 421-439, 2002.
- [131] Y. Lin and W. J. Zhang, "Aesthetic design for automobile interiors: critical and conceptual framework," in *2004 IEEE International Conference on Systems, Man and Cybernetics*, The Hague, Netherlands, pp. 6313-6317, 2004
- [132] A. Sa-ngasoonsong and S. T. S. Bukkapatnam, "Willingness-to-pay prediction based on empirical mode decomposition," in *Industrial Engineering Research Conference*, Reno, NV, 2011.
- [133] W. McManus, "The link between gasoline prices and vehicle sales," *Business Economics*, vol. 42, pp. 53-60, 2007.
- [134] A. R. Mian and A. Sufi, "Household leverage and the recession of 2007 to 2009," *IMF Economic Review*, vol. 58(1), pp. 74-117, 2010.
- [135] B. Brühl, *et al.*, "A sales forecast model for the German automobile market based on time series analysis and data mining methods," in *Advances in Data Mining. Applications and Theoretical Aspects*. vol. 5633, P. Perner, Ed., Berlin: Springer , pp. 146-160, 2009.
- [136] S. Shahabuddin, "Forecasting automobile sales," *Management Research News*, vol. 32, pp. 670-672, 2009.
- [137] F.-K. Wang, *et al.*, "Using adaptive network-based fuzzy inference system to forecast automobile sales," *Expert Systems with Applications*, vol. 38, pp. 10587-10593, 2011.

- [138] E. Zivot and J. Wang, "Modeling financial time series with S-PLUS." vol. 191, New York, NY: Springer, 2006.
- [139] G. Adomavicius, *et al.*, "Modeling supply-side dynamics of IT components, products, and infrastructure: An empirical analysis using vector autoregression," *Information Systems Research*, vol. 23(2), pp. 397-471, 2012.
- [140] S. Das, "Modelling money, price and output in India: A vector autoregressive and moving average (VARMA) approach," *Applied Economics*, vol. 35, pp. 1219-1225, 2003.
- [141] J. Berkovec, "Forecasting automobile demand using disaggregate choice models," *Transportation Research Part B: Methodological*, vol. 19, pp. 315-329, 1985.
- [142] P. H. Franses, "Modeling new product sales: An applicaiton of cointegration analysis," *International Journal of Research in Marketing*, vol. 11, pp. 491-502, 1994.
- [143] A. Greenspan and D. Cohen, "Motor vehicle stocks, scrappage, and sales," *The Review of Economics and Statistics*, vol. 81, pp. 369-383, 1999.
- [144] C. A. Lave and K. Train, "A disaggregate model of auto-type choice," *Transportation Research Part A: General*, vol. 13, pp. 1-9, 1979.
- [145] F. L. Mannering and K. Train, "Recent directions in automobile demand modeling," *Transportation Research Part B: Methodological*, vol. 19, pp. 265-274, 1985.
- [146] S. Shahabuddin, "Forecasting automobile sales," *Management Research News*, vol. 32, pp. 670-672, 2009.
- [147] F.-K. Wang, *et al.*, "Using adaptive network-based fuzzy inference system to forecast automobile sales," *Expert Systems with Applications*, vol. 38, pp. 10587-10593, 2011.
- [148] M. G. Dekimpe and D. M. Hanssens, "The persistence of marketing effects on sales," *Marketing Science*, vol. 14, pp. 1-21, 1995.
- [149] J. R. Landwehr, *et al.*, "Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts," *Marketing Science*, vol. 30, pp. 416-429, 2011.

- [150] M. G. Dekimpe, *et al.*, "Long-run effects of price promotions in scanner markets," *Journal of Econometrics*, vol. 89, pp. 269-291, 1998.
- [151] P. Rothman, "Forecasting asymmetric unemployment rates," *The Review of Economics and Statistics*, vol. 80, pp. 164-168, 1998.
- [152] C. Cheng, *et al.*, "Time series forecasting for nonlinear and nonstationary processes: A review and comparative study," *IIE Transaction* (Forthcoming).
- [153] Z. Kong, *et al.*, "Nonlinear sequential Bayesian analysis-based decision making for end-point detection of chemical mechanical planarization (CMP) processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24(4), pp. 523-532, 2011.

APPENDIX

Table A1: The information criteria for the selection of lag length (ADF test)

Lag	log (Y ₁) (p* = 5)		log (Y ₂) (p* = 4)		log (X ₁) (p* = 3)		log (X ₂) (p* = 6)	
	AIC	SBC	AIC	SBC	AIC	SBC	AIC	SBC
1	-1.5265	-1.4975	-2.0035	-1.9746	-2.3756	-2.3466	-9.1985	-9.1696
2	-1.5616	-1.5229	-2.0471	-2.0084	-2.3750	-2.3363	-9.1990	-9.1603
3	-1.5662	-1.5178	-2.0555	-2.0070	-2.3757	-2.3272	-9.2041	-9.1557
4	-1.5723	-1.5141	-2.0797	-2.0215	-2.3818	-2.3236	-9.2060	-9.1477
5	-1.5728	-1.5047	-2.0783	-2.0102	-2.3749	-2.3068	-9.2032	-9.1351
6	-1.5713	-1.4934	-2.0848	-2.0069	-2.3682	-2.2903	-9.2073	-9.1293

Lag	log (X ₃) (p* = 5)		log (X ₄) (p* = 6)	
	AIC	SBC	AIC	SBC
1	-3.4436	-3.4146	-4.3233	-4.2943
2	-3.5318	-3.4932	-4.3789	-4.3402
3	-3.5381	-3.4896	-4.4411	-4.3927
4	-3.5390	-3.4807	-4.4479	4.3897
5	-3.5587	-3.4906	-4.4689	-4.4009
6	-3.5532	-3.4753	-4.4767	-4.3988

Table A2: Park's G Test for Stationarity (Original Variables)

Original Variables		Test for Stationarity with G(p,q)						
		G(1,2)	G(1,3)	G(1,4)	G(1,5)	G(1,6)	G(1,7)	G(1,8)
log (Y ₁)	χ ² -value	0.395	10.813	13.756	13.900	14.064	15.770	15.861
	P-value	0.530*	0.004	0.003	0.008	0.015	0.015	0.026
log (Y ₂)	χ ² -value	9.446	9.962	12.127	12.708	13.910	13.945	14.796
	P-value	0.039	0.030	0.016	0.013	0.007	0.007	0.002
log (X ₁)	χ ² -value	0.847	7.579	11.298	11.587	11.720	14.648	14.651
	P-value	0.357*	0.022	0.010	0.021	0.039	0.023	0.041
log (X ₂)	χ ² -value	4.196	9.778	9.778	10.205	10.230	14.778	15.322
	P-value	0.041	0.008	0.021	0.037	0.069	0.022	0.032
log (X ₃)	χ ² -value	7.549	12.542	13.636	14.189	14.574	15.691	16.018
	P-value	0.006	0.002	0.003	0.007	0.012	0.016	0.025
log (X ₄)	χ ² -value	3.404	8.941	10.196	10.269	11.710	12.073	12.686
	P-value	0.065	0.011	0.017	0.036	0.039	0.060	0.080

* The null hypothesis cannot be rejected at 10% significance level.

Table A3: Park's G Test for Stationarity (First Differenced Variables)

First Differenced Variables		Test for Stationarity with G(p,q)						
		G(1,2)	G(1,3)	G(1,4)	G(1,5)	G(1,6)	G(1,7)	G(1,8)
$\Delta\log(Y_1)$	χ^2 -value	0.141	1.076	3.909	3.963	6.801	7.389	7.425
	p-value	0.707	0.585	0.271	0.411	0.236	0.286	0.386
$\Delta\log(Y_2)$	χ^2 -value	0.202	0.204	0.259	1.108	1.911	1.919	3.513
	p-value	0.653	0.903	0.968	0.893	0.861	0.927	0.834
$\Delta\log(X_1)$	χ^2 -value	0.537	3.302	4.922	4.922	10.128	10.440	10.556
	p-value	0.464	0.192	0.178	0.295	0.072*	0.107	0.159
$\Delta\log(X_2)$	χ^2 -value	0.050	0.136	4.152	4.307	6.568	7.126	13.152
	p-value	0.822	0.934	0.245	0.366	0.255	0.309	0.068
$\Delta\log(X_3)$	χ^2 -value	0.496	0.663	0.975	1.114	1.120	2.551	4.961
	p-value	0.481	0.718	0.807	0.892	0.952	0.863	0.665
$\Delta\log(X_4)$	χ^2 -value	2.959	4.598	4.835	4.838	6.033	12.942	12.974
	p-value	0.085*	0.100	0.184	0.304	0.303	0.044	0.073

* The null hypothesis is rejected at 10% significance level.

Table A4: Lag Length Selection Criteria

Lag	AICC	HQC	AIC	SBC	FPEC
1	29.2137	29.2359	29.2135	29.2702	4.867E12
2	29.1034	29.1403	29.1029	29.1976	4.357E12
3	29.0663	29.1177	29.0653	29.1980	4.196E12
4	29.0440	29.1098	29.0422	29.2132	4.101E12
5	29.0470	29.1270	29.0443	29.2537	4.109E12
6	29.0607	29.1547	29.0568	29.3047	4.161E12

Table A5: Overall Model Criteria

MODEL	Overall Model Criteria				
	AIC	SBC	AICC	HQC	FPEC
VAR(2)	-28.6001	-28.2906	-28.5971	-28.4778	3.79E-13
VARX(2,2)	-29.3493	-28.8076	-29.3399	-29.1351	1.79E-13
VECM(2)	-28.6509	-28.3806	-28.6486	-28.544	3.61E-13
VECMX(2,2)	-29.3829	-28.8799	-29.3748	-29.184	1.73E-13

Table A6: Modeling Details

Model	Model Equation
Quadratic Trend Model	$\log(Y_{1t}) = \alpha + \beta_1 t + \beta_2 t^2 + \varepsilon_t$
IMA	$\Delta\log(Y_1)_t = \varepsilon_t + \beta_1 \varepsilon_{t-1}$
ARIMA	$\Delta\log(Y_1)_t = a_1 \Delta\log(Y_1)_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$
ADL	$\Delta\log(Y_1)_t = \sum_{i=1}^{p=4} a_i \Delta\log(Y_1)_{t-i} + \sum_{i=0}^{n=1} b_i \Delta\log(Y_2)_{t-i} + \sum_{i=1}^{n=2} c_i \Delta\log(X_1)_{t-i} + \varepsilon_t$
ARIMAX	$\Delta\log(Y_1)_t = \sum_{i=1}^{p=1} a_i \Delta\log(Y_1)_{t-i} + \sum_{i=0}^{n=2} b_i \Delta\log(Y_2)_{t-i} + \sum_{i=0}^{n=0} c_i \Delta\log(X_1)_{t-i} + \sum_{i=1}^{q=1} d_i \varepsilon_{t-i}$

VITA

Akkarapol Sa-ngasoongsong

Candidate for the Degree of

Doctor of Philosophy

Dissertation: FORECASTING FOR NONLINEAR AND NONSTATIONARY
SYSTEMS USING INTRINSIC FUNCTIONAL DECOMPOSITION
MODELS

Major Field: Industrial Engineering and Management

Biography:

Personal Data: Born in Bangkok, Thailand on June 22, 1981

Education:

Received the B.Eng. degree in Electrical Engineering from Mahidol University, Nakhon Pathom, Thailand in May 2003

Received the M.S. degrees in Industrial Engineering and Management, and in Agricultural Economics from Oklahoma State University in December 2009 and May 2010, respectively

Completed the requirements for the Doctor of Philosophy in Industrial Engineering and Management at Oklahoma State University in December 2014

Experience:

Akkarapol has research interests in econometrics, analytics and data mining. He has work experience as a research and development engineer at LG Electronics (Thailand), and failure and reliability analysis engineer at Hitachi Global Storage Technologies (Thailand). During his Ph.D. study, he has worked as a graduate research and teaching assistant for the School of Industrial Engineering and Management at Oklahoma State University (OSU).

Professional Memberships:

- Institute for Operations Research and the Management Sciences (INFORMS)
- The Industrial Engineering Honor Society (Alpha Pi Mu)
- Institute of Electrical and Electronics Engineers (IEEE)
- Institute for Industrial Engineers (IIE)