The Pennsylvania State University

The Graduate School

College of the Liberal Arts

SEMI-NONPARAMETRIC DISCRETE EVENT FORECASTING IN ECONOMICS AND FINANCE

A Thesis in

Economics

by

Guang Guo

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2004

The thesis of Guang Guo was reviewed and approved* by the following:

Herman J. Bierens
Professor of Economics
Thesis Advisor
Chair of Committee

N. Edward Coulson
Professor of Economics

Philip A. Klein
Professor Emeritus of Economics

Coenraad Pinkse
Associate Professor of Economics

Timothy Simin
Assistant Professor of Finance

Robert C. Marshall
Professor of Economics
Head of the Department of Economics

*Signatures are on file in the Graduate School

ABSTRACT

Probabilistic forecasts play a significant role in a wide variety of economics activities. However, established econometric modeling approaches for probabilistic forecasts may yield unsatisfactory forecasting performance due to model misspecification. In my dissertation, I try to minimize such a risk by introducing a semi-nonparametric modeling approach for probabilistic forecasting. The new approach combines the ARMA memory index modeling approach of Bierens (1988) with the semi-nonparametric estimation method that uses wavelet basis to construct a flexible functional form. With this combination, we are able to avoid imposing restrictive constraints on the specification of critical components of conditional probability functions, i.e., the lag structure and distribution functions of error terms. As a result, it is possible that the new modeling approach will lead to improved forecasting performance if the reduction of modeling bias is significant. To test the usefulness of the new approach, we compare the relative performance between the new modeling approach and traditional forecasting models in both Monte Carlo experiments and real-world applications including business cycle regime forecasting and the forecast of rank performance of stock returns. The experimental and empirical results suggest that the new modeling approach can outperform traditional modeling approaches due to the flexibility of the model specification and the way various nonlinearities in the dependence of conditional probabilities on information variables are captured by ARMA memory indices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The probability of the completion of this thesis would be zero a.s. if I had made this journey alone. It is the result of five years of work whereby I have been accompanied and supported by many people. It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them.

The first person I would like to thank is my supervisor Herman J. Bierens. The idea of this dissertation originated from the summer project ‘Turning Point Analysis’, which I started to work on with him during Summer 1999. Not only did he provide a motivating, enthusiastic, and critical atmosphere during the many discussions we had to help me not to get lost during the development of the thesis, he also gave generous financial supports to my research activities and personal life during each summer session in the past five years. Besides, his own success story has always been a major force that inspires me to stick with my dream and efforts. He could not even realize how much I have learned from him. In addition to the guidance on my thesis, Herman was also as close as a good friend to me and gave precious advice on everything that will bring me more opportunities of a smooth career path and successful life. I never doubt that it is my best luck in my life to have Herman J. Bierens as my thesis advisor.

Many thanks go to Professor Phillips A. Klein, a distinguished expert in Business Cycle Studies. The discussion between us has clarified many technical issues concerning the Business Cycle Turning Point forecasts and facilitated the application of Discrete-Event-Forecasting model to the Business Cycle regime forecasts.

I would also like to thank Professor Timothy Simin, who let me recognize the potential utility of Discrete-Event-Forecasting model in the field of financial studies. It is his critical comments that made it possible that my research in stock return forecasting could make contribution to the finance literature.

I would also like to thank the other members of my PhD committee who monitored my work and took effort in reading and providing me with valuable comments on earlier versions of this thesis: N. Edward Coulson, Coenraad Pinkse. I thank you all.

Special thanks goes to Misty Crawford who has made magical efforts to help me with my thesis writing. It is her advice that made me to deliver my thinking and idea more clearly and effectively in English.

My colleagues of my department all gave me the feeling of being at home at work and exchanged great ideas during our talk on my thesis research. Subramaniam Chandrasekhar, Yong Dai, Vijay R. Krishna, Hung-Pin Lai, Yih-Ming Lin, Abhiroop Mukhopadhyay, Jun Xue, Chun-Chieh Wang and Li Wang, many thanks for being your colleague.

I feel a deep sense of gratitude for my father and mother who grant their maximal supports to my overseas study during their post-retirement period when they need assistance in life most. In addition, their ideologies of serving the society unselfishly and hard-working spirit always provides a persistent inspiration for my journey in all my life. I would also express my thanks to my two brothers, Haiteng Guo and Yang Guo, who has released me from the major concern over my parents family by taking care of them in these years.

I am very grateful for my wife Xiaofeng Wang for her love, patience and tolerance during the PhD period. As an ambitious woman who also possesses a remarkable passion for life, she has made considerable personal sacrifice in her career and life to accompany me through this period. One of the best experiences that we lived through in this period was the birth of our son Max C. Guo, who provided an additional and joyful dimension to our life mission.

Guang Guo

State College, Pennsylvaina

April, 2004

Chapter 1

PROBABLISTIC FORECASTING IN ECONOMICS

1.1 Probabilistic Forecasting and Its Link with Moment Forecasting

In terms of forecasting outcomes, economic forecasts can be classified into two types, moment forecasting and distribution forecasting. In moment forecasting, forecasters predict a particular moment of a random variable. A typical example is level forecast, i.e., $E(Y_t|\mathcal{F}_{t-1})$, where $E(\bullet|\mathcal{F}_{t-1})$ is the conditional mean and $\mathcal{F}_{t-1}$is the information set up to the period $t-1$. In distribution forecast, forecasters predict the whole probability distribution function of a random variable, i.e., $\Pr(Y_t < y|\mathcal{F}_{t-1})$ for $y \in \mathbb{R}$ (see Tay&Wallis 2000). Probabilistic forecast is a special case of distribution forecast, in which forecasters predict the probability of outcomes of a particular random event variable.

In certain cases, moment forecast can be used to do probabilistic forecast of a random event if the underlying data generating process of the event has a simple structure. For example, the random event variable X_t is defined as

$$X_t \quad \begin{cases} = 1 & \text{if } Y_t > v \\ = 0 & \text{otherwise} \end{cases}$$

where the stationary process $Y_t \tilde{} N(\mu_t, \sigma_t)$. Conditional mean forecasting and conditional volatility forecasting of Y_t will provide sufficient information to do probabilistic forecast of X_t. In some other cases, probabilistic forecast is needed to complement moment forecast if the forecaster wants to give a better description of potential uncertainty. However, if the underlying data generating process for the random event has non-standard stochastic regularities, the above connections between moment forecast and probabilistic forecast will be weakened. In that case, we need to choose specific forecasting models and methods for each type of forecasting.

1.2 Probabilistic Forecasting in Economic Studies

Probabilistic forecasting are particularly important if the outcomes of the target random event variables have significant impacts on a variety of economic activities. We provide 3 examples. First, regime switches of business cycles have deep influences on all aspects of the economy (see Niemira & Klein 1994 , Lahiri & Moore 1991). Consequently, probabilistic forecasting of business cycle turning points is of tremendous interest to various parties. Second, the outcome of the Federal Reserve Bank's policy decision draws wide public attention due to its strong implication on the future economy. Improved probabilistic forecasting of those outcomes will help investors and business managers make better investing decisions (see Hamilton & Jorda 2002). Third, in the financial market, the event that a stock falls into a particular performance category will change the follow-up performance of this stock dramatically (see Jagadeesh and Titman 1993). Accurate probabilistic forecasting on such an event will provide portfolio management with substantive benefits.

When the forecasted discrete events mainly concern the behavior of individual economic agents, we have, in econometric literature, well-established methods to set up econometric models, i.e., the discrete-choice model. But if the event we are studying concerns more complicated components such as the dynamic of the whole economy or the state shift of the financial markets, the theoretic backup for the appropriate econometric modeling approach is seriously lacking. We should take great caution in choosing the analysis tools that will serve our forecasting purpose best. In the next section, we will introduce the main methodologies for probabilistic forecasting in the economics literature.

1.3 Key Methodologies in Probabilistic Forecasting

The econometric methodologies in probabilistic forecasting can be divided into two groups, those belonging to the Bayes approach and those to the time series approach.

1.3.1 Bayes Approach

The Bayes approach assumes that before a particular event takes place, there is a structural change in the stochastic behavior of some observable information variables, known as leading indicator series. In order to do probabilistic forecasting of the target event, we need to find a way to detect the structural change of the leading indicator series.

Classic Composite Leading Indicator Method

In the classic Composite Leading Indicator method, we first choose a leading indicator series. After combining them into a composite leading indicator, we adopt rules of the thumb to predict the onset of a particular event. (see Hymans 1973) For example, in business cycle turning point forecasting, we have the '3CD' forecasting rule. When the composite leading indicator declines in three consecutive quarters, a recession is considered imminent and an alarm of downturn is triggered.

This method implicitly assumes that the distribution of the Composite Leading Indicator series has the following regularity. When the state of the economy is good, it is impossible to observe 3CD pattern on this series. Only when the state of the economy gets "bad", 3CD becomes a feasible event. More exactly,

$$\Pr(X_t = 1|S_t = 1) = 0$$

and

$$\Pr(X_t = 1|S_t = 0) > 0$$

where S_t is the state variable for the whole economy and the event variable $X_t = 1$ if 3CD happens and $X_t = 0$ otherwise. As a result,

$$
\begin{aligned}
\Pr(S_t &= 0|X_t = 1) \\
&= \frac{\Pr(X_t = 1|S_t = 0)\Pr(S_t = 0)}{\Pr(X_t = 1|S_t = 0)\Pr(S_t = 0) + \Pr(X_t = 1|S_t = 1)\Pr(S_t = 1)} \\
&= 1
\end{aligned}
$$

. That is, if we observe 3CD, we will infer that the state of economy has become bad ($S_t = 0$) and trigger the downturn alert.

One peculiar strength of such a method is its flexibility. The rule of thumb can be constructed based on various hypotheses. This is important when there is no established theoretic work to explain the cause to the target event and the event is rare. One obvious weakness of this method is that subjective factors often play a role.

'Sequential Probability Recursion' Scheme

The Sequential Probability Recursion scheme was first proposed in Neftici (1982). In this method, an explicit Bayes inference method replaces the rule of the thumb. In addition, a regime switch structure is formally imposed on the data-generating-process of the composite leading indicator.

The main content of the regime switch framework is that the distribution of the composite leading indicator will shift between two different settings. For example, let $\{Y_t\}_{t=1}^{\infty}$ be the first difference of the composite leading indicator.

$$Y_t = \mu_0 + \varepsilon_{0,t} \text{ if } S_t = 0;$$

$$Y_t = \mu_1 + \varepsilon_{1,t} \text{ if } S_t = 1$$

with constant $\mu_0 << \mu_1$ and $\varepsilon_{0,t}$ $\varepsilon_{1,t}$ are i.i.d sequences that follows a Gaussian process. The density distribution of $\varepsilon_{0,t}$ and $\varepsilon_{1,t}$ are $f^0(\bullet)$ and $f^1(\bullet)$, respectively. The shift cannot be observed at the time it occurs; however, we can use the Bayes rule to detect it. The probability to estimate is $\pi_k = \Pr(Z \leq k \mid Y_{k,} Y_{k-1}, ...)$ where Z is the random variable indicating the date of the regime switches .

The dynamic Bayes inference scheme is

$$\pi_{k+1} = \frac{[\pi_k + \Gamma_k(1-\pi_k)]f^1(y_{k+1})}{[\pi_k + \Gamma_k(1-\pi_k)]f^1(y_{k+1}) + (1-\pi_k)[1-\Gamma_k]f^0(y_{k+1})} \tag{1.1}$$

where $\Gamma_k = \Pr(z = k+1 | z > k)$.

Compared with the classic composite leading indicator method, this forecasting scheme yields a probabilistic forecast with more rich quantitative content. In addition, this scheme is able to accumulate information under the optimal stopping rule. Such a feature makes it a breakthrough in the Bayes approach of probabilistic forecasting.

Though the 'Sequential Probability Recursion' scheme adds more scientific rigor to probabilistic forecasting, there are some serious problems with this method. Although the regime switch model assumes that the regime switch is unobservable at the time it occurs, these regime changes have to be observable ex post in order for the forecasting scheme (1.1) to be implemented. In his paper, Neftici groups observations according to the decision of the NBER dating committee and estimates Γ_k, $f^0(\bullet)$ and $f^1(\bullet)$ using an empirical frequency method. The ad hoc grouping of the composite leading indicator series will trigger some problems. First, it introduces judgmental factors from the dating committee. Second, there is a potential goodness-of-fit bias in the dating process. The regime dating of the composite leading indicator always needs to be consistent with the regime dating of business cycles. Under this circumstance, there is an upward bias for the goodness of the model fit.

Hidden Markov Chain Model

The hidden Markov chain model (see Hamilton 1989) falls into the category of the Bayes approach since we rely on the Bayes rule to infer the likelihood of latent regimes from the behavior of observable variables. Unlike the 'Sequential Probability Recursion' Scheme, however, in the hidden Markov chain model, the law of motion for the latent regime variable and the distribution function of error terms are not ad hoc. Rather they are specified as part of the model. For example, let $\{Y_t\}_{t=1}^{\infty}$ be the first difference of the composite leading indicator and S_t be the latent binary event variable. Y_t is specified as

$$Y_t = \mu_1 + \psi_1(L)Y_{t-1} + \varepsilon_{1,t} \text{ if } S_t = 1;$$

$$Y_t = \mu_0 + \psi_0(L)Y_{t-1} \ + \varepsilon_{0,t} \text{ if } S_t = 0$$

where constant $\mu_0 << \mu_1$ and $\varepsilon_{0,t},\ \varepsilon_{1,t}$ are normally distributed i.i.d sequences.

The latent regime process, S_t, follows a first-order Markov chain with the transition matrix

$$\Pi = \begin{bmatrix} p^{00} & p^{01} \\ p^{10} & p^{11} \end{bmatrix}$$

where $p^{ij} = \Pr(S_t = j \mid S_{t-1} = i)$. A time-varying transitional probability can be added for a more realistic framework (see Filardo 1994). Because the distribution of Y_t is jointly determined by S_t and the history of Y_t, we can infer the value of S_t and construct the conditional likelihood function of Y_t using the historic information of Y_t through the Bayes rules.

The Model Misspecification Problem in the Bayes Approach

In each of these three methods which use the Bayes approach, there is improvement on the statistic rigor. From the classic method to 'Sequential Probability Recursion' Scheme, people add a regime switch structure. From 'Sequential Probability Recursion' Scheme to the hidden Markov chain model, a latent dynamic structure is added. However, in terms of the event forecasting performance, the new model is not necessarily better. Diebold and Rudebusch (1989) attempted to compare the classic composite leading indicator method with the 'Sequential Probability Recursion' scheme. Though the new model can provide more refined information on the event probability, it does not improve the number of correct predictions of turning points. As to the benefit of using the hidden Markov chain model, studies such as Filardo(1994), Hamilton and Perez-Quiros (1996) show that the predictive power of the leading indicator series is limited to only one or two months. One potential problem with the Bayes approach is the model misspecification of the structure of the latent event series. Such a mistake, however, will affect forecasting performance most because the dynamics of the event process plays a central role in determining the odds of each event outcome. A more flexible setting of the interdependence between the history of information variables and a future event's likelihood would help to alleviate this problem.

1.3.2 Time Series Approach

The time series approach does not impose a regime switch structure. The basic rationale for this approach is that past shocks to information variables have a persistent impact on the likelihood of the target event. If such an impact is stationary, then a time series model can be used to capture it.

Stochastic Simulation Method

In this method, each forecasted event is inherently associated with a continuous-distributed random variable. An example of this type of event is the variable E_t where $E_t = 1$ if a series drops by more than 5% in the next period and rises by more than 5% in the period after next. Normally, a series can be modeled as covariance-stationery process, e.g., an ARMA process and the forecasting procedure always starts with the estimation of the underlying time series model.

Now we give an example for illustration, which is adopted from Wecker (1979).

Let $\{Y_t\}_{t=1}^{\infty}$ be a time series that follow an AR(1) process, i.e.

$$Y_t = \alpha Y_{t-1} + \varepsilon_t \tag{1.2}$$

where $\varepsilon_t \sim N(0, \sigma^2)$.

Let X_t be

$$\begin{aligned} X_t &= 1, \text{ if } Y_t > 0, Y_{t-1} > 0, Y_{t-2} < 0, Y_{t-3} < 0, \\ &= 0, \text{otherwise.} \end{aligned} \tag{1.3}$$

Finally, define a target event Z_t as

$$Z_t = k$$

if we wait k periods before the next time $X = 1$.

To do the probabilistic forecasting, we need carry out a two-step procedure.

1. Estimate the data-generating-process, Y_t according to the specification as in (1.2), acquiring the estimator of α and σ.

2. With the estimator of α and σ, we do simulation to compute the probability of the

events of interest as follows. First, conditioning on observations of Y_t up to time t, we generate a series of artificial future values $y_{t+1}^{(n)}, y_{t+2}^{(n)}, y_{t+3}^{(n)}, y_{t+4}^{(n)}, ...$,according to (1.2). Then, we obtain the sequence $x_{t+1}^{(n)}, x_{t+2}^{(n)}, x_{t+3}^{(n)}, x_{t+4}^{(n)}, ...$ according to (1.3). Finally, we acquire the realization of $z_t^{(n)}$. After repeating this for N times, we can compute the empirical distribution of the event variable Z_t. We use a simulation method instead of analytic results because the latter involve a high dimensional integration, which might be infeasible in certain cases. For example,

$$\Pr(Z_t = 2) = \int I(\varepsilon_{t+2} > \alpha^2 Y_t + \alpha\varepsilon_{t+1}, \varepsilon_{t+1} > \alpha Y_t) dF(\varepsilon_{t+1}) dF(\varepsilon_{t+2})$$

Simulation will be a much faster way to obtain a probability estimator. It certainly lead to efficiency loss since a simulation procedure will involve sampling errors. As long as the sample size N is large enough, however, the variance of sampling error will approach zero and the accuracy is acceptable.

The major shortcoming of this approach is that in this approach we assume that we can correctly specify the underlying data-generating-process of the event variable. One crucial component of the model specification is the conditional distribution function of error items in the stochastic process. In order to implement probabilistic forecast, such specification should be as precise as possible. However, in standard econometric practices, the restriction on error items is always required to be as small as possible due to insufficient knowledge. As a result, we are subject to a great danger of model misspecification in the stochastic simulation approach.

Probit and Logit Model

The probit and Logit model also belong to the time series approach (see Estrella, A, & Mishkin F.S. (1998)). Unlike the stochastic simulation method, these two models require an direct functional relationship between the conditioning variables and forecasted event variables. In the Probit model case,

$$\Pr(X_t = 1) = \Phi(\alpha + \beta Y_{t-1}) \tag{1.4}$$

where $\Phi(\bullet)$ is an accumulative standard normal distribution function. In the Logit model case

$$\Pr(X_t = 1) = \frac{1}{1 + \exp(-\alpha - \beta Y_{t-1})} \tag{1.5}$$

Both Probit and Logit models were originally developed to study discrete-choice behavior in microeconomics. The functional forms of (1.4) and (1.5) are derived based on the specification of latent utility or profit function. If we apply them to other economic studies, i.e., macroeconomics studies, however, the modeling foundation will not be as strong . In those cases, great caution should be taken in model selections.

The Model Misspecification Problem in Time Series Approach

Like in the Bayes approach, the model misspecification constitutes a serious problem for the time series approach. Unlike in the Bayes approach, however, it is the specification of the error terms that cause the most concern. The impact of misspecification problems is the most serious in the stochastic simulation method because the link between model estimation and forecasting purpose is indirect. Using Probit and Logit models to do event forecasting can help alleviate this problem to some extent because both models construct the conditional probability function in a direct way. But if the departure from the true model is considerable, the impact of model misspecification will be substantial.

1.3.3 Comparison Between Two Basic Approaches and Potential Ways of Improvement

In the Bayes approach, the predictability of target events comes from the assumption that the data-generating-process of the leading indicator variables will switch between different stochastic settings before these events take place. The logic of this assumption is natural. It is not until abnormal events occur that a shift in the current state is suspected. In the Bayes approach, however, there is no specification on the time lag between the target event and the structural change of the leading indicator variables. In addition, the likelihood of the structural change of the leading indicators depends on the entire history of the conditioning variables due to the nonlinear inference scheme. Such a forecasting approach fits the situation in which the events to forecast have a significant influence but occur at a low frequency.

Models in the time series approach, however, impose a stationary correlation between the likelihood of the forecasted event and the conditioning variables. This approach is appropriate if it is obvious that the inter-dependence between the event variable and conditioning variables can be characterized explicitly in a lead-lag fashion.

The common problem shared by these two approaches is the vulnerability to model misspecification. In the Bayes approach, the existence and characteristics of the law of motion of the latent regime process is the most fragile part. For example, in the business cycle turning points forecasting literature, studies demonstrate that although a regime switch structure is significant for the key macro variables like GNP, it is not obvious that the leading indicator series also possess such a structure (see Filardo 1994, Layton 1996). In the time series approach, the weakest link is error term specification. When the forecasted event is associated with extreme behavior of future shocks and the high-order moment of the distribution of shocks varies significantly with the history of information variables, the standard model specification for moment forecast will fare poorly.

We consider a new approach which combines the features of these two approaches and tries to make improvement. Following the time series approach, we model the conditional probability explicitly and estimate the model in such a manner that helps to minimize the forecasting errors. To avoid serious misspecification, we construct the conditional probability function without specifying the particular distribution function. Following the Bayes approach, we allow the conditional probability to vary with the entire history of the information variables in an extremely flexible way. As a result, the dependence between conditioning variables and the target events can be highly nonlinear and the impact of past shocks can be spread over a considerably long timespan.

Chapter 2

DISCRETE-EVENT-FORECASTING MODEL

2.1 Basic Requirements for the New Model

2.1.1 A Direct Probability Estimator

In our new approach, we require that the forecasting model yield the probability forecast of the future events directly. By doing so, we can set up a direct link between the model estimation and the optimization of the probabilistic forecasting performance. Such a requirement adds obvious improvement to the forecast when the target event covers multi-period situations. More often that not, in multi-period forecasting, the model is estimated to optimize one-period-ahead forecast but is used to do multi-period forecast in practice. The impact of model misspecification will then become more remarkable since the error will be accumulated over time. Among all the methods discussed above, only the probit and logit model meet this requirement.

2.1.2 A Flexible Conditioning Structure

Most studies using the probit and logit models adopt functional forms that combine a linear model with a standard normal distribution function as in (1.4) or a logit function as in (1.5).

Such a structure is appropriate if we apply probabilistic forecasting in the study of discrete choice behavior. Seldom do we add dynamic consideration in those studies. Also specification of the error terms usually plays a minor role in determining the forecasting performance. In macro economy or financial market forecasting, however, a simple structure may not able to characterize the underlying data generating process as well as in microeconomics studies.

In our new approach, we would like to adopt a flexible way of modeling. First, we do not want to set stringent restrictions on the covariance structure in order to avoid serious model misspecification. Special events such as regime switches or extraordinarily strong performance of a stock cannot be attributable to a single sudden exogenous shock. The trigger is likely to have developed over a long horizon. Cutting off on the lag without sufficient justification will induce severe bias in the model estimation. Second, we do not want to set restrictions on the distribution function of error terms. For extreme events in macro economies and financial markets, the shocks that lead to these events usually do not behave regularly. For example, high frequency shocks in the stock market have fat tails and may be highly skewed (see Cootner 1964).

2.2 Setup of the Discrete-Event-Forecasting Model

The core part of our new forecasting model is to set up a direct and flexible link between the history of conditioning variables and the forecasted events. We use the following four-step procedure to implement such a setup. We denote the new modeling approach as Discrete-Event-Forecasting model.

a) Conditioning variables are clipped so that they become rationally-valued. For example, let $\{Y_t\}_{t=1}^{\infty}$ be the original information variable that we want to condition. We clip the data to acquire a new series $\{Y_t'\}_{t=1}^{\infty}$ through the following procedure.

$$Y_t' = a_i \text{ if } Y_t \in [b_i, b_{i+1}),$$

$i = 1, 2, ..., N.$

b) The entire history information of the discrete-valued new variable is mapped onto the real line to obtain a new variable. We call this variable an ARMA memory index of $\{Y_t', Y_{t-1}', Y_{t-2}'..., \}$.

For example, we define the ARMA memory index $M_t(\alpha)$ as

$$M_t(\alpha) = (1-\alpha)\sum_{i=1}^{\infty} \alpha^{i-1} Y'_{t-i} \tag{2.1}$$

c) A conditional probability function is set up to map the ARMA memory index into a probability measure for a specific event. For example, let $\{Y_t\}_{t=1}^{\infty}$ be a series of binary event variable . The forecasting model is

$$\begin{aligned} \Pr(X_t &= 1|Y'_{t-1}, ...) = \Pr(X_t = 1|M_t(\alpha)) \\ &= f(M_t(\alpha), \Theta) \end{aligned} \tag{2.2}$$

where $f(M_t(\alpha), \Theta)$ is a flexible functional form on $[0, 1]$ and Θ is the parameter set. In order to achieve the flexibility, we adopt a semi-nonparametric estimation method with a wavelet basis, i.e.,

$$f(M_t(\alpha), \Theta) = \sum_{k=K_0}^{K_1} c_{j_0,k}\varphi_{j_0,k}(M_t(\alpha)) + \sum_{j=j_0}^{j_1}\sum_{k} \theta_{j,k}\psi_{j,k}(M_t(\alpha))$$

where $\Theta = \{c_{j_0,k}, k = K_0, ..., K_1, \theta_{j,k}, j = j_0, ...j_1, k = K_1, ..., K_2\}$.

d) We use Maximum Likelihood Method to estimate Θ,i.e.,

$$\hat{\Theta} = \max_{\Theta} \widehat{L}(\Theta) \tag{2.3}$$

$$= \max_{\Theta} \prod_{t=1}^{n} \log\left[X_t f(M_t(\alpha), \Theta) + (1 - X_t)\left(1 - f\left(M_t(\alpha), \Theta\right)\right)\right] \tag{2.4}$$

In the following we will briefly address the main purpose of each step in our new approach.

First, the main goal of the data clipping procedure is to transform the conditioning information set $\mathcal{F}_t$ into a countable set. Intuitively, this procedure bundles some elements of the original information set together, condenses them into a single element and puts it into a new set. By doing this, we shrink the size of the conditioning information set.

Second, the shrinking of the original conditioning information set makes it possible for us to find a one-to-one mapping between the historic information of the conditioning variable and a real-valued random variable. With such a mapping, we can construct a one-dimension information variable, $M_t(\alpha)$, that contains all the information embedded in a high-dimensional information vector, $\{Y_t', Y_{t-1}', ...\}$.

Third, setting up the conditional probability function as in (2.2) is such that our new forecasting model possess all the above mentioned desirable properties. We have a direct probability estimator for the event we want to predict, i.e., $f(M_t(\alpha), \Theta) = \Pr(X_t = 1|Y_{t-1}', ...)$. The functional form is flexibly structured due to a semi-nonparametric estimation approach. Moreover, we adopt a wavelet basis to make the function estimation be sensitive to a local change in the functional shape.

2.3 Main Trade-off between the Discrete-Event-Forecasting Model and Traditional Modelling Approaches

The Discrete-Event-Forecasting model can help avoid major misspecification in the model structure of the conditional probability functions. This is particularly important when we choose events that are difficult to predict due to the complexity of the driving force. More often than not, the occurrence of these events is driven by shocks whose impacts diffuse slowly initially. Once these impacts become a public knowledge, the speed of diffusion will pick up and induce dramatic changes. For this reason, the dependence structure for the conditional probability function will cover a long horizon and the marginal impact of the changes in conditioning variable will be non-constant. Our new modelling approach can pick up these irregularities and improve probabilistic forecasting.

We must pay a price for these potential improvements. In the Discrete-Event-Forecasting model, we clip data to implement the ARMA memory index model. By doing so, we shrink the conditioning information set and increase the variance of forecasting error. Moreover, when the learning sample is small, we can not model our conditional probability function to such an extent that all the variation in the functional form can be captured. For that reason, the forecasting error may be enlarged further.

More formally, we define$\mathcal{F}_{t-1}$ as the $\sigma-$algebra generated by $\{Y_{t-1}, Y_{t-2}...\}$ and $\mathcal{F}_{t-1}^*$as the $\sigma-$ algebra generated by the clipped version, $\{Y_{t-1}', Y_{t-2}'...\}$. Note that $\mathcal{F}_{t-1} \subset \mathcal{F}_{t-1}^*$

When the conditional function $E(X_t|\mathcal{F}_{t-1}) = Z_t^1$, $E(Y_t|\mathcal{F}_{t-1}^*) = Z_t^*$ are both correctly specified, we have

$$\begin{aligned} MSE_o &= E\left[Y_t - E(Y_t|\mathcal{F}_{t-1})\right]^2 \\ &= E\left[Y_t - Z_t^1\right]^2 \end{aligned}$$

and by the law of iterated expectation,

$$\begin{aligned} MSE_c &= E\left[Y_t - Z_t^*\right]^2 \\ &= E\left[Y_t - Z_t^1\right]^2 + E\left[Z_t^* - Z_t^1\right]^2 \end{aligned}$$

When the conditional function $E(Y_t|\mathcal{F}_{t-1})$ is misspecified, say, $Z_t^0 \neq E(X_t|\mathcal{F}_{t-1})$ but Z_t^0 is still $\mathcal{F}_{t-1}$ measurable, by the law of iterated expectations,

$$\begin{aligned} MSE_m &= E\left[Y_t - Z_t^0\right]^2 \\ &= E\left[Y_t - Z_t^1\right]^2 + E\left[Z_t^1 - Z_t^0\right]^2 \end{aligned}$$

We define $Dif_{d,f}$ as follows.

$$\begin{aligned} Dif_{d,f} &= MSE_m - MSE_c \\ &= E\left[Z_t^1 - Z_t^0\right]^2 - E\left[Z_t^1 - Z_t^*\right]^2 \end{aligned}$$

$E\left[Z_t^1 - Z_t^*\right]^2$ is the variance increase due to the information set shrinkage while $E\left[Z_t^1 - Z_t^0\right]^2$ is the variance increase due to model misspecification. The sign of $Dif_{d,f}$ will determine which is a better forecasting method.

2.4 Two Important Econometrics Issues for the Discrete-Event-Forecasting Model

2.4.1 ARMA Memory Index Model

Basic Content

One core methodology for the discrete events forecasting model is the ARMA memory index model developed in Bierens (1988). A distinctive feature of this modelling approach is that it allows a flexible way of modelling the covariance structure in a conditional expectation function. As shown by Bierens (1988) , under mild regularity conditions, conditioning on the entire history of a rational-valued time series process is equivalent to conditioning on a single random variable which captures all the information contained in the past of the process. This random variable is called ARMA memory index since it can be formed as an autoregressive moving average of past observations. For example, let Y_t' be a univariate rational-valued process. Under mild conditions, using the ARMA memory index approach, we have

$$E(X_t|Y_{t-1}', Y_{t-2,}', ...,) = E(X_t|M_t(\alpha)) \tag{2.5}$$

where $E(\bullet|\bullet)$ is a conditional expectation function and $M_t(\alpha)$ follows (2.1) with $\alpha \in (-1, 1)\backslash S$, where S is a countable subset of $(-1, 1)$. The conditional expectation function, which conditions on the entire past of the time series involved, now becomes a Borel measurable function of a single ARMA memory index. Modelling the original function is impossible due to the 'curse of dimensionality' problem. Modelling the transformed function, $E(X_t|M_t(\alpha))$, however, can be implemented given that we can choose a flexible nonlinear functional form.

The proof of these theorems on ARMA memory index model is complicated (see Bierens 1988). But the main idea is intuitive. The key is the existence of a one-to-one mapping between the real line and a infinite-dimensional rational-valued vector. Due to such a mapping, the conditional probability can be transformed into a function of the index without information loss. Next, we give an example of such a mapping.

Let $Y_j', j = 0, 1, 2...$be a sequence of variables taking values in the set $\{0, 1, 2, ...k\}$. Let for $\alpha \in (0, 1)$,

$$M(\alpha) = (1-\alpha)\sum_{j=0}^{\infty}\alpha^j Y_j'$$

and let $\mathcal{M}_\alpha$ be the set of all possible values of $M(\alpha)$.

If $Y_0' = s \in \{0, 1, 2, ...k\}$, then

$$\begin{aligned} M(\alpha) &= (1-\alpha)s + \alpha(1-\alpha)\sum_{j=0}^{\infty}\alpha^j Y_j' \geq (1-\alpha)s \\ M(\alpha) &\leq (1-\alpha)s + \alpha(1-\alpha)\sum_{j=0}^{\infty}\alpha^j k \\ &= (1-\alpha)(s + \frac{\alpha}{1-\alpha}k) \end{aligned}$$

If

$$\alpha < \frac{1}{k+1} \tag{2.6}$$

then the intervals $[(1-\alpha)s, (1-\alpha)(s + \frac{\alpha}{1-\alpha}k)]$ are disjoint, hence for every $M \in \mathcal{M}_\alpha$,we can determine the value of Y_0' as follows:

$Y_0' = s$ iff $M \in [(1-\alpha)s, (1-\alpha)(s + \frac{\alpha}{1-\alpha}k)]$

Once the value of Y_0' has been established, define

$$M^{(1)} = \frac{M - (1-\alpha)Y_0'}{\alpha}$$

Then

$$Y_1' = s \text{ iff } M_1 \in [(1-\alpha)s, (1-\alpha)(s + \frac{\alpha}{1-\alpha}k)]$$

More generally, for $Y_t', t > 1$

$$Y_t' = s \text{ iff } M^{(t)} \in [(1-\alpha)s, (1-\alpha)(s + \frac{\alpha}{1-\alpha}k)]$$

where

$$M^{(t)} = \frac{M^{(t-1)} - (1-\alpha)Y'_{t-1}}{\alpha}$$

With the above algorithm, it is easy to show that the mapping between M and $\{Y'_t\}_{t=0}^{\infty}$ is one-to-one.

We should keep in mind that (2.6) is a sufficient but not necessary condition. In general, this can be done for any $\alpha \in (-1,1) \setminus S$ where S is a subset of $(-1,1)$ with Lebesgue measure 0.

Strength of ARMA Memory Index Approach in Economics

The ARMA memory index model will be particularly useful when the following situation exists. First, there is nonlinearity in the dependence structure for the underlying data-generating process. Only when we combine the historic observations of information variables together can we find a significant impact from the past. Second, the shock that promotes the chance of target events diffuse its influence slowly. Therefore, movements of information variables in the distant past will still have a significant impact on the likelihood of the target event.

The above situations will take place when the target event is not driven by a simple exogenous shock. Extreme events in Macro-economies and financial markets usually fall into this category. The main cause of these events is a combination of both human factors and technology factors. A simple dependence structure between conditioning variables and events to be forecasted does not exist. Hints towards the likelihood of target events are scattered in the history of conditioning variables. Those hints will have little predictive power if we consider them separately. The ARMA memory index approach makes it possible to find useful patterns by allowing conditioning on the whole history of information variables in a flexible way

2.4.2 Semi-Nonparametric Estimation with Wavelet Basis

A crucial factor in the success of ARMA memory index model is modelling a flexible functional form for the transformed conditional probability function of the ARMA memory index, i.e., (2.5). The latter can be fulfilled through the second core methodology for the discrete-events-forecasting model, the semi-nonparametric estimation method. In semi-nonparametric estimation, an unknown functional form with little prior knowledge imposed is estimated. Therefore, the object to be estimated is the entire structure of a functional form rather than an unknown scalar or vector as we see in standard parametric estimation.

The basic rationale behind semi-nonparametric estimation can be summarized by the following observation. Due to the isomorphic property of Hilbert Space (see appendix for definition) to l^2 space, a*ny infinite dimensional real vector can be decomposed as the 'sum' of its projection on the basis vectors; any functional form in a separable inner product space can be decomposed as 'sum' of orthogonal basis component (Fourier Series).* Therefore, estimating a functional form becomes equivalent to finding a coefficient vector (Fourier Coefficients) for components of an orthogonal basis. As a result, it is possible to apply the standard parametric estimation methods into functional estimation (see appendix Al).

Fourier Analysis The main purpose of the Fourier analysis is to find a simple way to represent an arbitrary functional form. The following theorem sets up a natural link between Fourier analysis and basic properties of an orthonormal basis (see more in the appendix).

Theorem 1 *An inner product space is separable if and only if it has a complete orthonormal sequence (x_n).Furthermore, in a separable inner product space, any $x \in X$ can be written uniquely in the form $x = \sum_{n \in \mathbf{N}} \langle x, x_n \rangle x_n$ for any orthonormal basis (x_n).*

Under certain conditions, Fourier series will have good convergence properties for piecewise smooth functions and bounded-variation functions (see appendix A). Due to these properties, we can approximate any functional form in a large class of functional space through a linear combination of functional components in a Fourier series.

Wavelet Analysis In Fourier analysis, a functional form is broken into frequency components. Each frequency component in Fourier analysis is a periodic function. If a functional form has a cyclical curve pattern, then it only takes a small number of Fourier coefficients to obtain a good estimator of this pattern. The efficiency is due to the global property of Fourier analysis. But the global property also brings a fatal shortcoming. If a curve pattern is a local phenomenon for a functional form, then we must use a lot of components of Fourier series to cover it. The Fourier analysis will become extremely inefficient under such circumstances. A way to overcome this problem is to find a functional basis, whose components are localized in its time domain. That is, the function value of each component is far away from zero only on a small region of the component's domain. As a result, Fourier coefficients of this basis will give only local frequency information of a functional form. If such a basis can be constructed, then coefficients of this basis will give the local 'content' of a functional form in the time domain and frequency domain simultaneously, which is known as time-frequency analysis.

The wavelet analysis is introduced in order to implement this idea. The strength of wavelet basis in functional analysis relies on the time-frequency analysis. The ability of carrying out time-frequency analysis improves the sensitivity of functional estimation to drastic local behavior of a functional form. It ensures detection of an anomaly in a functional shape with a small number of parameters. If we consider such a feature in a dynamic way, the estimation method could absorb new information in the functional form much more quickly than a regular one.

Brief History of Wavelets

The general form of wavelet is sequence $\{\psi_{j,k}(x)\ j,k \in Z\}$,s.t. $\psi_{j,k;} = 2^{-\frac{j}{2}}\psi(2^{-j}x - k)$ constitutes an orthonormal basis for the space $\mathbb{L}^2(\mathbb{R})$,i.e.,

$$\int \psi_{j,k}(x)\psi_{i,m}(x)dx \begin{cases} = 0 \text{ if } i \neq j \text{ or } k \neq m \\ = 1 \qquad\qquad \text{otherwise} \end{cases}$$

Such a formation contains two types of operation, the dilation such as $\psi_j(x) = \psi(2^{-j}x)$ and the translation such as $\psi_k(x) = \psi(x + k)$.

The first wavelet to be discussed in the literature is Haar Wavelet, which is a compactly supported wavelet. It was worked out by Haar in 1909. He tried to find an orthonormal base which can uniformly approximate any continuous function on a compact support. Opposite to the Haar wavelets, the Littlewood-Paley wavelets have compact supports in Fourier transforms. The first wavelet with excellent analytic properties in the time-frequency domain is due to Stromberg(1982). The wavelet is C^k with an exponential decaying speed in its functional form. Later, Meyer built up a wavelet with a compact support (see Meyer(1985)). He found the wavelet while he was trying to establish a proof of nonexistence of such a basis. Immediately after that, the first example of biorthogonal basis was constructed by Tchamitchian (1987). In the following year, Battle (1987) and Lemeir (1988) independently constructed the wavelets with the same structure, which is called Battle&Lemeir wavelets. Due to the lack of systematic theoretic foundation behind the constructions, these results seem more like miraculous incidences. Finally, Multi-resolution Analysis (MRA), first introduced by Mallat (1988) based on his knowledge of image processing, offered a unified framework to explain the success of all these cases. It is MRA that demonstrates that the wavelet is a synthesis of ideas originated in recent development in physics, engineering and pure mathematics.

Multi-resolution Analysis and General Approach of Constructing Wavelets

The Multi-resolution analysis has its root in the subband coding and quadrature mirror filtering(see Meyer 1993, Chapter 3), which are popular in image processing. The image of each object can be analyzed in a hierarchical way. It can be first decomposed into two orthogonal parts, a trend component W_0 and a fluctuation component V_0. The trend component shows a smooth and regular sketch of an original image, while the fluctuation component is the deviation from the regular sketch. After that, the trend component is left intact. The fluctuation component, however, is further analyzed into another pair of a trend component and fluctuation component, W_1 and V_1. With such a decomposition procedure, the images remaining in trend components will be increasingly volatile. The size of fluctuation components V_j, however, shrink to be negligible. Since trend components in the later stages are capable of characterizing more volatile variation of a functional form, such analysis is called a high resolution analysis. Correspondingly, the analysis in early stages is low resolution analysis. Multi-Resolution Analysis (MRA) is the mixed usage of them. According to the construction, the two components in each step are orthogonal to each other, i.e.,$W_j \perp V_j$. Since each trend component in a stage is a part of the fluctuation component of the previous stage, trend components are themselves orthogonal to each other i.e., $W_j \perp W_{j'}$, $j \neq j'$. Therefore, in the end, an image is decomposed into an infinite number of trend components which are orthogonal to one another.

The link between wavelet basis $\{\psi_{j,k}(x)\ j, k \in Z\}$ and Multi-Resolution Analysis (MRA) is as follows. To construct a wavelet basis, we need to find a sequence of subspaces V_j such that such that $V_{j-1} \supset V_j \supset V_{j+1}$, $\cup_{j\in\mathbb{Z}} V_j = L^2(\mathbb{R})$ and $\cap_{j\in\mathbb{Z}} V_j = \{0\}$. For each V_j, we have an orthognormal basis, $\{\varphi_{j,k}(x),\ k \in Z\}$. From $\{\varphi_{j,k}(x)\}$, we can construct an orthognormal basis

$\{\psi_{j,k}(x)\ k \in Z\}$ for each subspace W_j, which is the orthogonal complement to V_j, i.e.,$V_{j-1} = V_j \bigoplus W_j$. Due to the properties of V_j and W_j, the system $\{\psi_{j,k}(x)\ \}_{j,k}$ will constitute an orthogonormal basis for $\mathbb{L}^2(\mathbb{R})$. Normally, we denote $\varphi_{j,k}(x)$ as the scaling function and $\psi_{j,k}(x)$ the mother wavelet function.

In constructing $\{\psi_{j,k}(x)\ k \in Z\}$ from $\varphi(x)$, we use two properties of $\psi_{j,k}(x)$. First, $\psi_{j,k}(x)$ belongs to V_{j-1} and we can have a Fourier representation of it as

$$\psi_{j,k}(x) = 2^{-\frac{(j-1)}{2}} \sum_n f_n \varphi(2^{-(j-1)}x - n)$$

where $\varphi(2^{-(j-1)}x - \bullet)$ is the orthogonal basis for V_{j-1}. Second, $\psi_{j,k}(x)$ is orthogonal to each basis function belonging to V_{j-1}, .i.e.,

$$2^{-\frac{(j-1)}{2}} \int \psi_{j,k}(x)\varphi(2^{-(j-1)}x - n) = 0 \text{ for all } n.$$

From those two conditions, we obtain one possible approach of constructing wavelet basis, that is,

$$\psi(x) = 2^{\frac{1}{2}} \sum_n (-1)^{n-1} h_{-n-1} \varphi(2x - n) \tag{2.7}$$

or

$$\hat{\psi}(\xi) = e^{i\xi/2} m_0(\frac{\xi}{2} + \pi)\hat{\varphi}(\frac{\xi}{2}) \tag{2.8}$$

where

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_n h_n e^{-in\xi}$$

is called low pass filter because of the following property

$$\hat{\varphi}(\xi) = m_0(\frac{\xi}{2})\hat{\varphi}(\frac{\xi}{2}) \tag{2.9}$$

or

$$\varphi(x) = 2^{\frac{1}{2}} \sum_n h_n \varphi(2x - n) \tag{2.10}$$

The function $m_0(\bullet)$ is able to filter out the high frequency information of a function form from its low frequency counterpart directly or it can construct a function with a large scale from its small scale counterpart through its Fourier coefficients h_n. Note that the method of construction is not unique. The equation (2.7) can also be replaced with

$$\psi(x) = \sum_n (-1)^n h_{-n+1} \varphi(2x - n) \tag{2.11}$$

We refer readers to the appendix for more formal illustration and related literature.

Example of Wavelets

Haar This is the simplest case of wavelets basis.

1.The scaling function $\varphi(\cdot)$ of Haar wavelet is

$$\begin{aligned} \varphi(x) &= 1 \text{ if } 0 \leq x < 1 \\ &= 0 \ \text{ otherwise.} \end{aligned}$$

2. Define $V_0 = \{f(x) : f \text{ is constant on the interval } [k-1, k),\ k \in Z \ \}$, of which the above function is an orthonormal basis.

3. Using (2.10), m_0 is

$$h_n = \sqrt{2} \int \varphi(x)\varphi(2x - n)dx \quad \begin{aligned} &= \tfrac{1}{\sqrt{2}} \ \text{ if } \ n = 0, 1 \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

4. Finally, using (2.11), we obtain

$$\begin{aligned} \psi(x) &= \sum_n (-1)^{n-1} h_{-n-1} \varphi(2x - n) \\ &= \frac{1}{\sqrt{2}}\varphi(2x) - \frac{1}{\sqrt{2}}\varphi(2x - 1) \end{aligned} \tag{2.12}$$

The Haar wavelet is compactly supported. but it is discontinuous. The consequence is that it has bad frequency localization. It is inappropriate to use such a function basis to approximate a function with reasonable smoothness, even though it converges everywhere.

Meyer Wavelet 1.The Fourier transform of the scaling function is

$$\begin{aligned}\hat{\varphi}(x) &= \sqrt{2\pi} & |\xi| \leq 2\pi/3 \\ &= \sqrt{2\pi}\cos[\frac{\pi}{2}v(\frac{3}{4\pi}|\xi|-1)],\ 2\pi/3 \leq & |\xi| \leq 4\pi/3 \\ &= 0 & \text{otherwise}\end{aligned}$$

where v is a smooth function satisfying

(a)

$$\begin{aligned}v(x) &= 1 \text{ if } x > 0 \\ &= 0 \text{ otherwise}\end{aligned}$$

(b)

$$v(x) + v(1-x) = 1$$

2. Define V_0 as the space spanned by this set. Then V_j is a closed subspace spanned by the $\varphi_{j,k} = 2^{-\frac{j}{2}}\varphi(2^{-j}x - k)$.

3. The m_0 is

$$m_0(\xi) = \sqrt{2\pi}\sum_{l \in Z}\hat{\varphi}(2(\xi + 2\pi l))$$

4. Finally, the wavelet basis is defined using (2.8)

$$
\begin{aligned}
\hat{\psi}(\xi) &= e^{i\xi/2}\left[m_0(\xi/2+\pi)\right]\hat{\varphi}(\xi/2) \qquad (2.13)\\
&= \sqrt{2\pi}e^{i\xi/2}\sum_{l\in Z}\hat{\varphi}(\xi+2\pi(2l+1))\hat{\varphi}(\xi/2)\\
&= \sqrt{2\pi}e^{i\xi/2}\left[\hat{\varphi}(\xi+2\pi)+\hat{\varphi}(\xi-2\pi)\right]\hat{\varphi}(\xi/2)
\end{aligned}
$$

Spline Wavelet This class of wavelet is also called the *Battle-Lemarie family wavelet*. It is also called spline wavelet, because the scaling function for is B-spline of degree n, $n \in \mathbb{N}$[1]. In fact, *Haar wavelet* is a special case of the spline wavelet with degree 0. The spline wavelet with degree 1 is called *Franklin Wavelet*.

1. The scaling function is piecewise linear function

$$
\begin{aligned}
\phi(x) &= 1-|x|\,, 0 \le x \le 1\\
&= 0 \qquad \text{otherwise}
\end{aligned}
$$

The Fourier transform of it is

$$
\hat{\varphi}(\xi) = (2\pi)^{-1/2}(\frac{\sin\xi/2}{\xi/2})^2
$$

By normalizing, we get

$$
\hat{\varphi}^{\#}(\xi) = (\frac{1+2\cos^2\xi/2}{3})^{-1/2}\hat{\varphi}(\xi)
$$

2. By (2.9), the corresponding $m_0^{\#}(\xi)$ is then

$$
m_0^{\#}(\xi) = \frac{\hat{\varphi}^{\#}(2\xi)}{\hat{\varphi}^{\#}(\xi)} = \cos^2(\xi/2)\left[\frac{1+2\cos^2(\xi/2)}{1+2\cos^2(\xi)}\right]^{1/2}
$$

3. Fourier transform of the mother wavelets then is

[1] **B-spline function**
The B-spline funtion of order n is the set of function f,such that the restriction of f to $(n, n+1]$ if polynomial of degree n.

$$\begin{aligned}
\hat{\psi}(\xi) &= e^{i\xi/2}\left[m_0^{\#}(\xi/2+\pi)\right]^* \hat{\varphi}^{\#}(\xi/2) \\
&= e^{i\xi/2}\sin^2(\xi/4)\left[\frac{1+2\sin^2(\xi/4)}{1+2\cos^2(\xi/2)}\right]^{1/2}\hat{\varphi}^{\#}(\xi/2)
\end{aligned} \tag{2.14}$$

Using the same logic, we can construct the *B-spline wavelet* with degree n for any $n \in \mathbf{N}$.The procedure is as follows:

1.The scaling function satisfies

$$\hat{\varphi}^n(\xi) = \frac{e^{-im\xi/2}(\frac{\sin\xi/2}{\xi/2})^n}{(F_m(e^{-i\xi}))^{1/2}}$$

where

$$F_m(x) = \sum_{-m+1}^{m-1} N_{2m}(m+k)x^k$$

$N_{2m}(m+k)$ can be calculated using following iterative algorithm

$$\begin{aligned}
N_2(k) &= \delta_{k,1} \\
N_{n+1}(k) &= \frac{k}{n}N_n(k) + \frac{n-k+1}{n}N_n(k-1)
\end{aligned}$$

2. The mother wavelet is computed as

$$\hat{\psi}^n(\xi) = -(\frac{4}{i\xi})^n e^{-i\xi/2}(\sin\xi/4)^{2n}\sqrt{\frac{F_n(-z)}{F_n(z)F_n(z)}} \tag{2.15}$$

where $z = e^{-iw/2}$. We refer interested readers for more detailed description of the above construction to Chui (1992).

Compactly Supported Wavelet (Daubechies Wavelets) All the above wavelets were constructed before the discovery of Multi-Resolution Analysis. MRA offers a unified framework to them. Daubechies Wavelets, however, rely heavily on Multi-Resolution Analysis and are good examples to demonstrate the strength of MRA. However, the compactly supported wavelet does not have an explicit expression form as the above examples. In order to plot the Daubechies wavelets, we need to use some special algorithm. Even so, the usage of Daubechies Wavelets in engineering fields has been overwhelming since it was invented. The major attraction comes from the compactness of its support. When we fix the level of resolution, as long as the estimated function has compact support, the number of wavelet components we use to approximate the functional form is always finite. This makes computation extremely fast. Since the way to construct Daubechies Wavelets is too complicated and is of little help for our later discussion, we skip this part. Interested readers could refer to Daubechies(1992) for detail.

Wavelets and Statistics

Functional estimation It is also called signal estimation. The main advantage of wavelet analysis in functional estimation comes from its capability to carry out localized multiscale decomposition. Although multiscale analysis is not new, the localization property distinguishes wavelet analysis from many traditional statistical analysis tools.

The main task of functional estimation is to effectively extract the functional form of a signal from contaminated observations. Usually, the signal has a structure that can be represented by a smooth function. Since a function may have a different level of smoothness on different scales, multiscale decomposition can help filter out the smoother structure. The main working mechanism of wavelet analysis on functional estimation can be explained by the following example due to Dohono et al (1995).

1. The regression model is

$$y_i = f(x_i) + \sigma\varepsilon_i, i = 1, 2, ...n$$

where x_i are equally spaced points and ε_i are i.i.d Gaussian noise.

2. Given that $f(x_i)$ satisfies some regularity conditions, it can be represented as

$$f(x_i) = \sum_k \alpha_k \varphi_{0,k}(x_i) + \sum_{j,k} \theta_{j,k}\psi_{j,k}(x_i)$$

where $\varphi_{0,k}(\cdot)$ is a scaling function and $\psi_{j,k}(\cdot)$ is a wavelet function. The coefficient are computed as

$$\tilde{\alpha}_k = \sum_i f(x_i)\tilde{\varphi}_{0,k}(x_i) \tag{2.16}$$

and

$$\tilde{\theta}_{j,k} = \sum_i f(x_i)\tilde{\psi}_{j,k}(x_i) \tag{2.17}$$

Now, we cannot observe $f(x_i)$ directly. To get an estimator of the above coefficients, we rely on its noised version y_i. Then the coefficients will be

$$\hat{\alpha}_k = \sum (f(x_i) + \delta\varepsilon_i)\tilde{\varphi}_{0,k}(x_i) = \tilde{\alpha}_k + \xi_k$$

and

$$\hat{\theta}_{j,k} = \sum (f(x_i) + \delta\varepsilon_i)\tilde{\psi}_{j,k}(x_i) = \tilde{\theta}_{j,k} + \xi_{j,k}$$

The $\xi_k, \xi_{j,k}$ are i.i.d Gaussian noise with distribution $N(0, \sigma_k^2)$ and $N(0, \sigma_{j,k}^2)$ respectively.

3. Using a thresholding method, we screen out a wavelet coefficient estimator which has very small value, that is,

$$\hat{\theta}_{j,k}^{(t)} = \hat{\theta}_{j,k} I(\hat{\theta}_{j,k} > \lambda),$$

where $I(\cdot)$ is the indicator function.

4. After that, we use the estimator $\hat{\alpha}_k$ and $\hat{\theta}_{j,k}^{(t)}$ to reconstruct the function $f(\cdot)$ by

$$f(x_i) = \sum_k \hat{\alpha}_k \varphi_{0,k}(x_i) + \sum_{j,k} \hat{\theta}_{j,k}^{(t)} \psi_{j,k}(x_i)$$

It should be noted that the threshold λ is a function of sample size n and variance factor σ. Since the latter is unobservable, we need to estimate it as we estimate the functional form for which an iterative algorithm can be used in many cases. Since coefficients will be shrunk if they are below thresholds, this method is called wavelet shrinkage method.

Density Estimation The second major area of statistical application of wavelet is density estimation. Again, using orthonormal series expansion to estimate a density function is by no means a new practice. There are two main approaches to implement the estimation technique.

(1) The first approach uses a representation theorem to carry out the semi-nonparametric

estimation This looks like a parametric estimation. An econometric application is Gallant and Tauchen (1987).

(2) The 2nd approach uses the idea of kernel estimation to carry out nonparametric estimation. The kernel is generated with a orthogonal series. As is shown in the appendix, the Dirichlet kernel function is the key to the good convergence properties of Fourier series. In particular,

$$\int f(x)D_n(y-x)dx \rightarrow f(y) \text{ as } n \rightarrow \infty$$

The $D_n(y-x)$ is called $\delta - sequence$ because its limit behaves like a δ function[2].

In wavelet density estimation, several approaches have been explored. A typical one has inherited the same idea as approach (2). Wavelet series are used to create a $\delta - sequence$. We set the reproducing kernel function of subspace V_0 in Multi-resolution Analysis as the starting point for the construction of a kernel function. A general formula is

$$q(x,y) = \sum_k \varphi(x-k)\varphi(y-k)$$

and

$$q_m(x,y) = 2^{2m} \sum_k \varphi(2^m x - k)\varphi(2^m y - k)$$

As $m \rightarrow \infty, q_m(x,y) \rightarrow \delta$ function.

Such a method is also called linear wavelet density estimator(see Gilbert 2001) Another group of estimator, nonlinear estimator, gains more interest.

The general form is

---

[2] **Dyadic-Delta Function:**

Dirac Delta Function is a function δ_x satisfying:

$$\delta_x(t) = 0 if t \neq x$$

and

$$\int \delta_x(t)dt = 1$$

Therefore, $\int f(t)\delta_x(t)dt = f(x)$ if $f(t)$ is well defined at point x.

$$f(x)=\sum_{k} c_{j_0,k}\varphi_{0,k}(x)+\sum_{j=j_0}^{j_1}\sum_{j,k}\theta_{j,k}\psi_{j,k}(x)$$

where

$$c_{j_0,k}=\frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0,k}(X_i) \tag{2.18}$$

$$\tilde{\theta}_{j,k}=\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(x_i) \tag{2.19}$$

As in Donoho et al.(1996), a simple non-linear wavelet estimator can be defined via thresholding

$$\hat{f}(x)=\sum_{k} c_{j_0,k}\varphi_{0,k}(x)+\sum_{j=j_0}^{j_1}\sum_{k}\hat{\theta}_{j,k}\psi_{j,k}(x) \tag{2.20}$$

where

$$\hat{\theta}_{j,k}=\delta^{h}(\theta_{j,k},\lambda)$$

For more information on wavelet and statistics and time series analysis, we refer interested readers to Antoniadis & Oppenheim(1995) and Percival (2000).

Combining the Semi-Nonparametric Estimation with Wavelet Analysis

Wavelet analysis is applied in a similar way as in Gallant and Nychka (1987) in this study. We adopt the series expansion procedure to represent the functional form for estimation. We then use Maximum Likelihood Estimation to estimate wavelet coefficients $c_{j_0,k},\psi_{j,k}$ in the representation form as

$$f(M_t(\alpha),\Theta)=\sum_{k=K_0}^{K_1} c_{j_0,k}\varphi_{j_0,k}(M_t(\alpha))+\sum_{j=j_0}^{j_1}\sum_{k}\theta_{j,k}\psi_{j,k}(M_t(\alpha))$$

where $\Theta=\{c_{j_0,k},k=K_0,...,K_1,\theta_{j,k},j=j_0,...j_1,k=K_1,...,K_2\}$. In order to obtain a consistent estimator of the functional form, we apply the method of sieve (see Grenander 1979, Geman & Hwang 1982). In this method, the number of the components chosen to construct the functional form is always a small fraction of data size. As a result, the variance of coefficient estimators can be kept small although we may face the bias in the functional form in a small sample case. As the sample size grows large, the bias of the functional form also shrinks due to series expansion. As long as wavelet coefficients of a functional form dampen quickly enough, an estimator will get arbitrarily close to the true functional form quickly.

The combination of wavelet basis and semi-nonparametric estimation makes it possible to combine the strength of both parametric estimation and non-parametric estimation. Non-parametric estimation is pure local analysis while parametric estimation is global. Local analysis can avoid serious bias in model estimation while global analysis can improve estimation efficiency if the local departure is minor. Applying semi-nonparametric estimation to the wavelet basis is a way to balance the bias and the variance in functional estimation.

Chapter 3

MONTE CARLO EXPERIMENT FOR DISCRETE-EVENT-FORECASTING MODEL

In this chapter, Monte Carlo experiments are designed to study the performance of the new forecasting model.

3.1 Experiment Design

3.1.1 Target Event

The Discrete-Event-Forecasting model is designed to do probabilistic forecasting of special economic events. The accuracy of such forecasts is sensitive to the model specification on the conditioning structure and the specification of error terms. In our experiment design, we will examine the potential improvement provided by the Discrete-Event-Forecasting model when the underlying process departs from a standard ARMA process significantly.

Let Y_t be a univariate stationary time series. We choose the following events to forecast.

(1) Single-period event.

(a) **Mild Gain Event:** the time series under study, Y_t, is higher than a value that is a little higher than the average at the period t.

$$X_{1,g,t} = I(Y_t > \bar{Y}_+)$$

where $\bar{Y}_+$ is the mild gain value. In our experiment, we choose $\bar{Y}_+$ such that 30% of the observed data is higher.

(b) **Severe Loss Event:** Y_t drops sharply at period t.

$$X_{1,l,t} = I(Y_t < \bar{Y}_-)$$

where $\bar{Y}_-$ is the severe loss value. In our experiment, we choose $\bar{Y}_-$ such that 5% of the observed data is lower.

(2) Multi-period event.

(a) **Mild Gain Event:** The time series under study, Y_t, is higher than a 1-period mild gain value at the period t, or higher than a 2-period mild gain value for the period t and $t+1$, or higher than a 3-period mild gain value for the period t, $t+1$ and $t+2$.

$$X_{m,g,t} = I(Y_t > \bar{Y}_{1,+} \text{ or } Y_t + Y_{t+1} > \bar{Y}_{2,+} \text{ or } Y_t + Y_{t+1} + Y_{t+2} > \bar{Y}_{3,+})$$

In our experiment, $\bar{Y}_{1,+} = \bar{Y}_+, \bar{Y}_{2,+} = 2 \times \bar{Y}_+, \bar{Y}_{3,+} = 3 \times \bar{Y}_+$.

(b)**Severe Loss Event:** Y_t drops heavily at t, or t and $t+1$, or t, $t+1$ and $t+2$.

$$X_{m,l,t} = I(Y_t < \bar{Y}_- \text{ or } Y_t + Y_{t+1} < \bar{Y}_- \text{ or } Y_t + Y_{t+1} + Y_{t+2} < \bar{Y}_-)$$

.

If we consider Y_t to be investment gain, the probabilities of type (a) events indicate the odds of breaking even. The probabilities of type (b) events show the chance of heavy loss during the forecast period. We choose these types of events to forecast for several reasons.

First, we choose mild gain values so that the mild gain event will not be a big surprise if it takes place. Normally, the likelihood of this event is mainly affected by the position of the conditional mean. We choose severe loss values such that their happenings will be a bad surprise.

Both conditional mean and higher-order conditional moments determine the likelihood of such events. As a result, model misspecification problems may affect the forecasting performance of these two events in different ways.

Second, we compare single-period events forecasts and multi-period events forecasts because forecasts of events covering longer time horizons are likely to be more sensitive to model misspecification. The error due to misspecification will be accumulated in multi-period forecast. In real-world practice, events like multi-period severe loss events play a significant role in determining the risk of a multi-period investment plan.

3.1.2 Data

We generate an original time series Y_t using two approaches.

ARMA Approach.

In this approach, a process is generated using an explicit difference equation. We adopt four types of specification, specified in such a way that the deviation from a standard AR process grows.

(1) AR(3) process

$$Y_t = \phi(L)Y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \tilde{} N(0, 1)$.

(2) ARMA(3,3) process

$$Y_t = \phi(L)Y_{t-1} + \psi(L)\varepsilon_t$$

where $\varepsilon_t \tilde{} N(0, 1)$.

(3)ARMA(3,3) process with mixed distribution with fixed weight

$$Y_t = \phi(L)Y_{t-1} + \psi(L)\varepsilon_t$$

where

$$\begin{aligned}\varepsilon_t &= \varepsilon_{1,t} \text{ with probability } p \\ &= \varepsilon_{2,t} \text{ with probability } 1-p\end{aligned}$$

where $\varepsilon_{1,t}\tilde{}N(0,1)$ and $\varepsilon_{2,t}\tilde{}N(0,\sigma^2)$, $\sigma > 1$.

(4) ARMA(3,3) process with mixed distribution with time-varying weight

$$Y_t = \phi(L)Y_{t-1} + \psi(L)\varepsilon_t$$

and

$$\begin{aligned}\varepsilon_t &= \varepsilon_{1,t} \text{ with probability } \frac{1}{1+\exp(-\varphi(L)Y_{t-1})} \\ &= \varepsilon_{2,t} \text{ with probability } \frac{\exp(-\varphi(L)Y_{t-1})}{1+\exp(-\varphi(L)Y_{t-1})}\end{aligned}$$

$\varepsilon_{1,t}\tilde{}N(0,1)$ and $\varepsilon_{2,t}\tilde{}N(0,\sigma^2)$, $\sigma > 1$.

For each setting , 1000 data points are generated. We use 600 data points to estimate the conditional probability function, 200 to cross validate and 200 to do out-of-sample forecasting. 50 samples for each type of specification are generated.

Non-Equation Approach.

In this approach, we do not follow an explicit difference equation to generate the data. Therefore, we ourselves do not have a clear idea of the exact description of the law of motion for the underlying data-generating process. We first generate a i.i.d series Z_t which follows a standard normal distribution, i.e., $Z_t\tilde{}N(0,1)$ and $cov(Z_{t_1}, Z_{t_2}) = 0, t_1 \neq t_2$. Then we generate Y_t as follows.

Let $V_t = \sum_{\tau=1}^{t} Z_t$,

If $t = 1$,

(1) if $(1-\alpha)Z_t > \bar{Z}$,then $Y_t = Z_t$ and $C_t = 1$;otherwise, $Y_t = \alpha Z_t, C_t = 0. P_1 = Y_1$;

If $t > 1$,

(1) if $|V_t - (P_{t-1} + Z_t)| > \bar{Z}$ and $[V_t - (P_{t-1} + Z_t)]\, Z_t > 0$, then $Y_t = [V_t - (P_{t-1} + Z_t)]\,(1 + \beta_1)$ and $C_t = 1$;

(2) if $|V_t - (P_{t-1} + Z_t)| > \bar{Z}$, $C_{t-1} = 1$ and $[(P_{t-1} + Z_t) - V_t]\, Z_t > 0$, then $Y_t = Z_t(1 + \beta_2)$ and $C_t = 0$;

(3) if $|V_t - (P_{t-1} + Z_t)| < \bar{Z}$, then $Y_t = \alpha Z_t$ and $C_t = 0$.

$P_t = \sum_{\tau=1}^{t} Y_\tau$;

We can consider Z_t as original shocks for an asset's price. V_t is the intrinsic value of the asset at period t and P_t is the price of the asset at the period t.

The price P_t evolves in the following way. If Z_t is small, its value will be reflected on P_t partially, i.e., αZ_t. When such under-valuation persists too long and $|V_t - (P_{t-1} + Z_t)|$ becomes too large, the price needs correction. However, there will be overcorrection each time, i.e., $[V_t - (P_{t-1} + Z_t)]\,(1 + \beta_1)$. Moreover, if Z_t has the same sign as $(P_{t-1} + Z_t) - V_t$ immediately after the correction is overdone, the overcorrection will be extended even further, i.e., $Z_t(1+\beta_2)$.

The data created through this non-equation approach will behave in a way that is more difficult for standard time series models to capture. As a result, the forecasting performance will be deteriorated more severely if model misspecification occurs. By doing so, we reduce bias in the model comparison.

3.1.3 Model Setup

Discrete-Event-Forecasting model

The conditional probability function is set up through the Discrete-Event-Forecasting model.

$$\Pr(X_{1,g,t} = 1 | Y'_{t-1}, Y'_{t-2}, ...) = F(M_t(\alpha)|\alpha)$$

where $X_{1,g,t}$ is the mild gain event variable and Y'_t is a clipping version of Y_t ,e.g.,

$$Y'_t = \sum_{i=1}^{p} a_i I(Y_t \in A_i)$$

where $\cup_{i=1}^{p} A_i = R$ and $A_i \cap A_j = \Phi$ and

$$M_t(\alpha) = (1-\alpha)\sum_{j=0}^{\infty} \alpha^j Y'_{t-1-j},$$

We specify $F(x|\alpha)$ such that $F(x|\alpha) \in [0,1]$ for all $x \in R$. It is constructed by letting

$$F(x|\alpha) = \frac{1}{1+\exp(-g(h(x)))}.$$

where

$$h(x) = \ln(\frac{x}{1-x}), x \in (0,1).$$

and $g(x|\alpha)$ is

$$g_{j_0,j_1}(x) = \sum_{k=K_0}^{K_1} c_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j=J_0}^{J_1}\sum_{k=K_{j,0}}^{K_{j,1}} \theta_{j,k}\psi_{j,k}(x)$$

where $\varphi_{j,k}(x)$ is scaling function and $\psi_{j,k}(x)$ is the mother wavelet function. In our model, we choose spline wavelets (see chapter 2) with degree 1 for our estimation.

When the data sample is not large, we must choose an appropriate dilation truncation J_0, J_1 and translation truncation K_0, K_1, $K_{j,0}, K_{j,1}$ for the scaling function, $\varphi_{j,k}(x)$, and the mother wavelet function , $\psi(\xi)$, respectively.

In our model, we choose the following specification:

(1) For the scaling function: $J_0 = 0$,and $K_0 = -1$. So we have three scaling function with 0 resolution level (see Figure 3.1).

(2) For the wavelet function: $J_1 = 1$ and

a) $K_{1,0} = -2, K_{1,1} = 1$ (see Figure 3.1)

b) $K_{2,0} = -5, K_{2,1} =, 4$ (see Figure 3.1)

As a result, we have three basic models:

Model (1) only has scaling functions and the number of functions is 3.

Model (2) combines scaling functions (scale 0) with mother wavelet functions (scale 0) and the number of functions is 7.

Model (3) combines scaling function (scale 0), mother wavelet (scale 0) and mother wavelet (scale 1), and the number of functions is 17.

For each data process, we have 30 different Discrete-Event-Forecasting models (10 values for α and three resolution level). We do model selection through cross validation; that is, we compare the model performance all the Discrete-Event-Forecasting models using cross-validation samples. We then choose the model with the best performance for the final-round out-of-sample forecast.

AR model.

We estimate each simulated data process with an AR(p) model. Based on the estimation result, we compute the conditional probability using either an explicit formula (single-period event) or the stochastic simulation method (multi-period event) (see Chapter 1). We do model selection through cross validation, that is, we estimate the AR(p) process using the learning sample with various lag specifications. We then choose the model which has the best mean square error performance in the cross-validation sample.

3.1.4 Model Evaluation

We compare the model performance in terms of the Quadratic Probability Score (QPS), i.e.,

$$QPS = T^{-1}\sum_{t=1}^{T}[X_{1,b,t+1} - \Pr(X_{1,b,t+1} = 1|Y_t, Y_{t-1}, ..., Y_1)]^2 \tag{3.1}$$

This is the standard performance measure for probabilistic forecast in economics literature (see Diebold and Rudebusch 1989). It gives a measure of the accuracy of the probabilistic forecast which is analogous to the Mean Square Error for level forecasting.

3.2 Experiment Result Analysis

3.2.1 Single-Period Forecast

In Table 3.1, statistics are presented comparing probabilistic forecast performance on a single-period event between the AR(p) model and the Discrete-Event-Forecast (DEF) model. The statistics include (1) the absolute and relative Quadratic Probability Score performance between the AR model and the Discrete-Event-Forecast model, (2) Wilcox statistics for relative performance, (3) the number of cases in which Discrete Event Forecast model outperforms (4) the average Mean-square error for the AR(p) models and (5) the average Quadratic Probability Score for the Discrete Event Forecast model. In Panel A, probabilistic forecast on a single-period mild gain event are compared. In Panel B, probabilistic forecast are compared for a single-period severe loss event.

For data generated by the ARMA approach, i.e., the first four types of process, the AR(p) outperforms Discrete-Event-Forecast (DEF hereafter) model significantly in both cases, confirmed by t-statistics (Col 3) and Wilcox statistics (Col 4). Moreover, across the four types of simulated processes, for both the AR(p) model and DEF model, there is little difference in the absolute and relative forecasting performance in terms of Quardratic Probability Score. The relative advantage of AR(p) model still maintains even if its model fit deteriorates (Col 6) when bias due to model misspecification increases.

There are two main reasons for the inferior performance of the new model in this case. First, although we try to generate deviation from standard ARMA process, we still have symmetric error terms in all four cases. Under this circumstance, the approximation from the AR(p) model with symmetric normal errors is very good assuming that the lag long enough. Such a good approximation assures satisfactory performance for the short-run forecast. Second, there is an overfitting problem for the new approach. Comparing Col (7) and Col (2), we can easily find large discrepancies between in-sample and out-of-sample forecasting performance. This problem may disadvantage the new forecasting model significantly.

With respect to the data generated by the non-equation approach, however, we get encouraging results. In particular, in mild gain event cases, the out-performance of our new forecasting model is dominant though for severe loss event the AR(p) model still does better. This outcomes shows that when the series under study demonstrates non-standard regularities, the standard forecasting model cannot easily capture these features adequately. In this situation, the competitive advantage of our new model will surface.

3.2.2 Multi-Period Forecast

In Table 3.2, we present statistics comparing probabilistic forecasting performance on a multi-period event between the AR(p) model and the Discrete-Event-Forecast (DEF) model. The statistics contained in this table are exactly the same as above. For this type of forecast, the advantage of DEF model starts to show off .

For data generated by the ARMA approach, we emphasize two main features. First, an obvious change is that probabilistic forecasting performance deteriorates for AR(p) model (col 1) when the deviation from AR process increases. In mild gain multi-period event forecast, the Quadratic Probability Score of the AR model is 0.21947 for ARMA process with time-varying non-normal error and 0.14174 for the AR process with a standard normal error. For the DEF model, however, the impact of the deviation is almost negligible (col 2). The Quadratic Probability Score changes little across the four types of model, e.g., for the mild gain case, 0.21001, 0.2005, 0.20441, 0.20685. Such a difference leads to improved relative performance for the DEF model when the model misspecification problem becomes more severe.

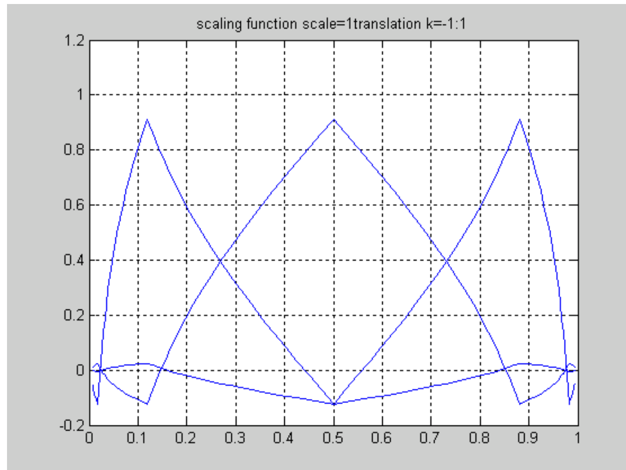
Second, the advantage of our new forecasting model is more significant in the severe loss event forecast, where higher order moments play an important role in determining the outcome. For the process with time-varying mixed normal error terms, the DEF becomes dominant in both Quadratic Probability Score (col 3) and the number of winning cases (col 5).

For the data generated by the non-equation approach, the advantage of our new forecasting model is further enhanced. The new model does significantly better in terms of Quadratic Probability Score in both mild gain event forecasts and severe loss event forecasts. In addition, it beats the AR model in the number of winning cases with significant margins in mild gain event forecasts and show strong competence in severe loss event forecasts.

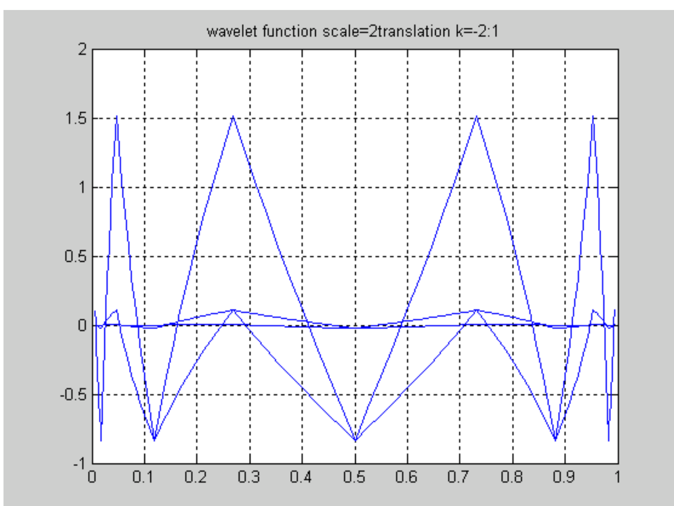
From these results, we can draw the following conclusion. In the multi-period forecasts, our new model provides strong competition to the traditional forecasting model. The cost of model misspecification for the AR(p) model becomes significant in forecasting multi-period events while our new model demonstrates an obvious advantage in its robustness to this problem. Such robustness demonstrates the advantage that our new model has in certain cases, verifying our original conjecture. Further research is under way to examine the relative advantage of our new forecasting model in different situations.

Figure 3.1: Plots of Scaling Functions and Mother Wavelets Functions for Franklin Wavelets

Scaling Functions with Scale 0



Mother Wavelet Functions with Scale 0



Mother Wavelet Functions with Scale 1

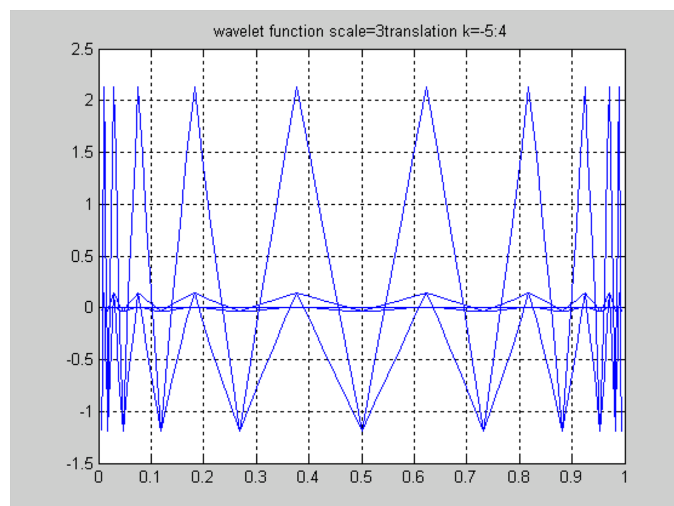


Table 3.1 Forecasting Performance Comparisons for Monte Carlo Experiments-- Single-Period Event Forecasts

In this table, statistics are presented comparing the AR (p) model and the DEF model in their probabilistic forecast performance on a single-period event. The statistics includes (1) the absolute and relative QPS performance between the AR model and the Discrete-Event-Forecast model (2) Wilcox statistics for relative performance (3) The number of winning cases for Discrete-Event-Forecasting model (4) Average Mean-square error for AR(p) model, and (5) Average QPS for the Discrete-Event-Forecast ing model in Cross Validation sample.

Panel A compares probabilistic forecast on a single-period break-even event.

Panel B compares probabilistic forecast on a single-period throat-cutting loss event

Panel A

Model	QPS_{AR}	QPS_{DEF}	QPS_{AR}-QPS_{DEF}	Wilcox-Statistics P-value4	# DEF win /Total	Average MSE (AR)	Average QPS (DEF)
AR(3)	0.11957	0.18641	-0.06684 (-8.6706)5	0	2/50	0.99559	0.16368
ARMA(3,3)_A1	0.11109	0.18404	-0.07296 (-10.8703)	0	1/50	1.3521	0.165
ARMA(3,3)_B2	0.11551	0.18906	-0.07355 (-11.7322)	0	0/50	1.724	0.16907
ARMA(3,3)_C3	0.10912	0.17617	-0.06705 (-9.6516)	0	0/50	3.0604	0.15398
Non-Equation	0.23357	0.21556	0.0180120 (7.6619)	0	40/50	4.2511	0.20736

Panel B

Model	QPS_{AR}	QPS_{DEF}	QPS_{AR}-QPS_{DEF}	Wilcox-Statistics P-value4	# DEF win /Total	Average MSE (AR)	Average QPS (DEF)
AR(3)	0.038525	0.069973	-0.03145 (3.8931)	0	3/50	1.0008	0.050507
ARMA(3,3)_A	0.030452	0.062416	-0.03196 (3.4499)	0	3/50	1.3521	0.03823
ARMA(3,3)_B	0.032143	0.067606	-0.03546 (3.3619)	0	3/50	1.715	0.044463
ARMA(3,3)_C	0.032938	0.061716	-0.02878 (3.7739)	0	3/50	3.0604	0.038776
Non-Equation	0.048459	0.059824	-0.01137 (5.7277)	0.000162	15/50	4.2511	0.039043

Note: 1. ARMA process with normal distributed error terms.

2. ARMA process with mixed normal distributed error terms. One normal distribution has variance 1 and the other have variance 2.25.The mixed weight is fixed. The probability for a standard normal error is 0.75 and the probability for the other one is 0.25.

3. ARMA process with mixed normal distributed error terms. One normal distribution has variance 1 and the other has variance 2.25.The mixed weight is time varying.

4. Null hypothesis is that the two models yield equal forecasting performance.

5. Value in brackets is the t-statistics.

Table 3.2 Forecasting Performance Comparisons for Monte Carlo Experiments--Multi-Period Event Forecasts

In this table, statistics are presented comparing the AR (p) model and the DEF model in their probabilistic forecast performance on a multi-period event. The statistics includes (1) the absolute and relative QPS performance between the AR model and the Discrete-Event-Forecast model (2) Wilcox statistics for relative performance (3) The number of winning cases for Discrete-Event-Forecasting model (4) Average Mean-square error for AR(p) model, and (5) Average QPS for the Discrete-Event-Forecast ing model in Cross Validation sample.

Panel A compares probabilistic forecast on a multi-period break-even event.

Panel B compares probabilistic forecast on a multi-period throat-cutting loss event.

Panel A

Model	QPS_{AR}	QPS_{DEF}	QPS_{AR}-QPS_{DEF}	Wilcox-Statistics P-value4	# DEF win /Total	Average MSE (AR)	Average QPS (DEF)
AR(3)	0.14174	0.21001	-0.06827 (4.7724)	0	2/50	0.99559	0.17947
ARMA(3,3)_A1	0.14071	0.2005	-0.05979 (4.9623)	0	1/50	1.3521	0.17969
ARMA(3,3)_B2	0.16089	0.20441	-0.04352 (5.5656)	0	0/50	1.724	0.18221
ARMA(3,3)_C3	0.21947	0.20685	0.012625 (7.5027)	0.29632	21/50	3.0604	0.16375
Non-Equation	0.31824	0.24135	0.076893 (9.3239)	0	39/50	4.2511	0.23511

Panel B

Model	QPS_{AR}	QPS_{DEF}	QPS_{AR}-QPS_{DEF}	Wilcox-Statistics P-value4	# DEF win /Total	Average MSE (AR)	Average QPS (DEF)
AR(3)	0.10704	0.14918	-0.04214 (5.0737)	0	2/50	1.0008	0.12847
ARMA(3,3)_A	0.11537	0.14342	-0.02806 (5.6878)	0	10/50	1.3521	0.11667
ARMA(3,3)_B	0.13049	0.13846	-0.00797 (6.664)	0.034456	20/50	1.715	0.11776
ARMA(3,3)_C	0.1874	0.124	0.063399 (10.6864)	0	41/50	3.0604	0.10029
Non-Equation	0.16139	0.13209	0.029304 (8.6398)	0.3516	24/50	4.2511	0.11279

Chapter 4

APPLICATION OF DISCRETE-EVENT-FORECASTING MODEL IN BUSINESS CYCLE REGIME FORECAST

In this section, we apply Discrete-Event-Forecasting model to carry out probabilistic forecast on business cycle regimes. We compare the Discrete-Event-Forecasting model with the Probit model in terms of both regime forecasting performance and turning points forecasting performance.

4.1 Literature Review

There have been three major improvements in regime forecast techniques since the introduction of the leading indicator approach in early 1940's.

First, Neftici (1982) developed 'Sequential Probability Recursion' scheme which provides a dynamic statistical inference method to identify structural changes in a stochastic data generating process. This scheme is able to predict forthcoming regimes by analyzing moving patterns of leading indicator series in a dynamic manner.

Second, Hamilton's latent regime switch model (Hamilton (1989)) solves the major deficiency of Neftici's algorithm. In Neftici's approach, one has to assume that structural changes of a composite leading indicator series can be observed ex post. In addition, one has to impose the assumption that the dating of regime switches in leading indicator series is independent from the dating of their counterparts in business cycles. Such an assumption is problematic because of the interdependence between the dating of business cycle turning points and the selection and dating of leading indicator series. Therefore, it is not surprising to have a satisfactory forecasting performance when we adopt the dating decision of NBER committee on both business cycles and leading indicator series. The latent regime switch model can avoid this problem by identifying the latent regime variable through a data-driven approach. Application of this model in dating turning points in leading indicator series for the purpose of business cycle regime forecast is well documented (see Lahiri and Wang 1994, Layton 1996).

Third, Watson and Stock (1991) use the unobservable component model to set up a more rigorous time series model framework for the selection of leading indicator series. This model assumes the existence of a latent common factor, which proxies the state of the economy. The dynamic structure of this factor is estimated using a Kalman Filter approach. Since the recession is defined as a specific pattern of the time path of the latent factor series, we can predict the recession by carrying out multi-period forecast of the latent factor. Unlike a regime switch model, the unobserved component model is still a linear model. The regime variable is not an exogenous variable. Its movement is highly dependent on the movement of the latent factor series.

These three models constitute three major steps in the development of econometric modelling techniques for business cycle regime forecasts. Variants of those models are numerous. Kling (1987) extends the stochastic simulation method of Wecker (1979) to forecast turning point on GDP within a VAR model framework. Koskinen and Oller (2000) develop a hidden-Markov-model (HMM) that adopts a Bayesian classification approach to classify the observations in order to carry out regime forecast. Kim and Nelson (1998) combines the regime switch model with the unobservable component model to create a more exquisite model structure that explores the latent regime-switch co-movement among major macro variables. Camacho and Perez-Quiros (2002) use a non-parametric time series approach to do regime forecast. Filardo

(1994) extends Hamilton's model with a time-varying transition matrix. Finally, Birchenhall, et al (1999) and Estrella and Mishkin (1998) adopt the logit and probit models to forecast regimes directly. For a comprehensive coverage on major techniques in business cycle regime and regime switches forecasting, we refer interested readers to Camacho and Perez-Quiros (2002).

These above developments add more statistical sophistication and rigor to regime identification and optimal prediction rules. They strengthened the scientific basis for regime forecasts. However, a general observation of the performance of these forecasting models is that the predictive power of information variable is reliable only in a short horizon (Estrella and Mishkin (1998) is an exception) . The effective forecasting lead is restricted to only one or two months ahead (see Filardo 1994, Hamilton and Perez-Quiros 1996).

4.2 Research Design

4.2.1 Basic Motivation

Our study is linked closely with Estrella and Mishkin (1998), which use a probit model to examine the predictive power of various financial variables in business cycle regime forecasts. The forecasted object in their study is the regime series created by NBER business cycle dating committee. In our study, we choose the same target. In addition, we will also choose the term spread of treasury bill interest rate, which is verified by Estrella and Mishkin as the most informative indicator series, as the conditioning variable.

The major distinction between our study and Estrella and Mishkin (1998) is the modeling approach. They adopt a standard probit model framework while we use the Dicrete-Event-Forecast model. The motivation of our choice is as follows.

First, in order to apply the probit model framework, the modeler has to assume that the regime process is determined by a latent variable Y_t^*. In addition, in most cases, forecasters will combine a linear structure with normal errors due to the concern of the tractability of statistic significance. Such a framework, which can be well justified in micro-economics studies might introduce deficiency in macro-economics studies due to model misspecification. In the latter, the existence of the latent variable Y_t^* is hard to be justified. Moreover, it is even harder to find an appropriate way of modelling the latent process when established theories for the driving force of regime switches are lacking. If the simple structure mentioned above contains serious bias in model specification, it will not be able to yield the optimal forecasting performance.

To conquer this problem, we use the Discrete-Event-Forecasting model instead to carry out probabilistic forecasting. This new modelling approach does not impose overt constraints on the specification of conditioning structures and error term distributions. The forecasting model structure is estimated through a data-driven approach. In case the model misspecification in standard models causes a large bias, the Discrete-Event-Forecasting model will bring forth a competitive advantage in terms of forecasting performance.

Second, in Estrella and Mishkin (1998), they use the level of term spread of treasury bill interest rate to predict future business cycle regimes. However, as to why the level of the term spread can shed light on the state of the economy, there is no conclusive explanation yet. One possible justification is that the level of the term spread actually indicates the trend of changes in term spread in recent periods, which will give hints on the regime of future economy. More specifically, a high level of term spread is usually accompanied with consecutive increase in the spread. This tendency might be linked with shocks that lead to subsequent regime switches. For this reason, we conjecture that if we use the change of the term spread as a conditioning variable and allow the business cycle shocks to diffuse their impact slowly over a long period, we might acquire a better forecasting performance. The Discrete-Event-Forecast model is an effective approach to implement such an idea.

4.2.2 Data

We use the Business Cycle chronology from National Bureau of Economic Research as the regime series to forecast. It is a monthly date series. For the conditioning variable, we created it based on the data series "FYGM3" and "FYGT10" from DRI database available through WRDS database from Wharton Business School of University of Pennsylvania. Both data series are monthly data. The "FYGM3" series is the Interest Rate of 3-month US Treasury Bills in Secondary Market. The interest rate is annualized rate and there is no seasonal adjustment. The "FYGT10" series is the Interest Rate of 10 year US Treasury with Constant Maturities. The interest rate is also annualized rate without seasonal adjustment. We transform these two dataset into the "Spread" series by computing the contemporary difference of them.

The data sample is divided into two parts: the in-sample learning part 1959.08-1996.12 (449 obs) which includes six peak turning points (1960.04, 1969.11, 1973.11, 1980.01, 1981.07, 1990.07) and out-of-sample part 1997.01-2001.12 (60 obs) which contains one peak turning point (2001.3) .

4.2.3 Model

Define S_t as the indicator series that indicate the US business cycle regime such that $S_t = 1$ if the economy expands and $S_t = 0$ if the economy takes a downturn.

Define Y_t as the financial variable that we think contains leading information for business cycle regimes.

Discrete-Event-Forecast Model

(1) We discretize the variable Y_t as

$$Y_t' = i, \text{if } \ t_{i-1} < Y_t' \leq \ t_i, i \geqslant 0$$

where $t_0 = -\infty, t_i < t_{i+1}$ and $t_I = +\infty$, where I is the total number of partition on the real line $\mathbb{R}$.

(2) With

$$Y_t^* = \frac{Y_t'}{I}$$

we restrict the variable into $[0, 1]$

(3) We create two ARMA memory indexes, that is,

$$\text{ARMA memory index I} : M_{y,1,t} = (1-\alpha) \sum_{i=1}^{+\infty} \alpha^{i-1} Y_{t-i}^*, \tag{4.1}$$

$$\text{ARMA memory index II} : M_{y,2,t} = (1-\lambda) \sum_{i=1}^{+\infty} \lambda^{i-1} (\triangle Y_{t-i}^*), \tag{4.2}$$

where $\triangle Y_t^*$ is the clipped version of $\triangle Y_t = Y_t - Y_{t-1}$.

(4) We model the conditional probability function as

$$\begin{aligned}\Pr(S_{t+k} &= 0|Y_t^*, Y_{t-1}^*, ...) \\ &= F\left[g(M_{y,t})\right] \\ &= F\left[g(h(\theta_1 M_{y,1,t} + (1-\theta_1)M_{y,2,t}))\right]\end{aligned} \tag{4.3}$$

where $F(\cdot)$ is modelled as

$$F(x) = \frac{1}{1+\exp(-g(h(x))}$$

with

$$h(x) = \ln(\frac{x}{1-x}), x \in (0,1).$$

and

$$g(x) = \sum_{k=-j}^{j} \beta_{1,}\varphi_{j_0,k}(x) + \sum_{i=2}^{3}\sum_{k=-j_i}^{j_i} \beta_{i,k}\psi(x)$$

For the Dicrete-Event-Forecasting model, we choose different resolution level for the wavelet basis. In this exercise, we choose three types, high(17 parameter) and medium (7 parameter) and low (3 parameter) (see chapter 3 for reference).

Probit Model

$$\begin{aligned}\Pr(S_{t+k} &= 1|\ Y_t, Y_{t-1}, ...) \\ &= \Pr(\varepsilon_{t+k} > \alpha_0 + \alpha_1 Y_t) \\ &= \Phi(\frac{\alpha_0 + \alpha_1 Y_t}{\sigma})\end{aligned}$$

where $\Phi(\bullet)$ is the accumulative distribution function of a standard normal distribution.

4.2.4 Model Evaluation

We set up two groups of performance measures to compare the probit model with the Discrete-Event-Forecasting model.

First, we use both the likelihood function (L) and Quadratic Probability Score (QPS) to measure the goodness of fit of forecasting models. We examine the forecasting performance for both in-sample and out-of-sample cases. The two measures are expressed as

$$L = T^{-1} \sum_{t=1}^{T} \log \left[I(S_{t+k} = 1) \Pr(S_{t+k} = 1|Y_t) + I(S_{t+k} = 0)(1 - \Pr(S_{t+k} = 1|Y_t)) \right] \tag{4.4}$$

and

$$QPS = T^{-1} \sum_{t=1}^{T} (S_{t+k} - \Pr(S_{t+k} = 1|Y_{t-1}))^2 \tag{4.5}$$

Second, we design a forecasting performance measure for the business cycle peaking point forecast. One major utility of business cycle regime forecasts is to predict the business cycle turning points, especially the peak turning points. In Estrella and Mishkin (1998), this issue has not been stressed. Because the major benefit of introducing the Discrete-Event-Forecasting model is to improve forecasting accuracy, we have to examine its performance in this aspect. This measure is defined as follows.

First, we set up the alarm threshold, say, 0.4. Then we apply the estimated conditional probability function $\Pr(S_{t+k}|Y_t, Y_{t-1,\ldots})$ to the data to carry out k-period-ahead forecasting. Once the conditional probability exceeds the alarm threshold, we make a call for a turning point. And we define four types of alarms.

1. *Prompt correct alarm.* The conditional probability exceeds the threshold before turning points occur. The date it predicts the turning points to happen is prior to or equal to the exact date of turning point. The alarm is maintained until the occurrence of the turning point.

More technically, we issue a correct prompt alarm at t^* with a positive lead $t' - t^*$ if

(1) $t' = \min\{t : s_t = 1, t^* \leq t \leq t''\}$, and

(2) $F(M_{y,t^*-k-1}) < P_{threshold}, F(M_{y,t-k}) > P_{threshold}$, for $t^* \leq t \leq t'$;

2. *Late correct alarm.* The conditional probability exceeds the threshold before turning points occur. But it indicates that a turning point will take place at a date later than the actual date of turning point.

Formally, we issue a late correct alarm at t^* with a negative lead $t^* - t'$ if

(1) $s_{t'-1} = 0$,and $s_t = 1$ for $t' \leq t \leq t''$, $s_{t''+1} = 0$

(2) $F(M_{y,t'-k}) < P_{threshold}$ and $t^* = \min\{t : F(M_{y,t-k}) > P_{threshold},\ t^{'} \leq t < t''\}$

3. *False alarm.* The conditional probability exceeds the threshold but falls below it before the actual arrival the turning point.

We issue a false alarm if

(1) $F(M_{y,t-k}) > P_{threshold}$, for $t^* < t < t'$, $F(M_{y,t'-k-1}) < P_{threshold}$

(2) $s_t = 0$ for $t^* \leq t \leq t'$.

4. *Missed alarm.* The conditional probability function does not issue a correct alarm for the turning point.

The performance of a forecasting model is measured in terms of the number of correct alarm, the number of false alarm, the number of missed alarm and the average lead or lag for the correct alarm.

4.3 Result

4.3.1 Table

Performance statistics are presented in two tables, table 4.1 and table 4.2. The table 4.1 demonstrates the in-sample performance while the table 4.2 shows the out-of-sample performance. The out-of-sample performance is critical to examine the quality of forecasting. However, the regime switches are rare incidents. In order to have a large learning sample to reduce the variance in the semi-nonparametric estimation, we split the whole sample so that only one turning point is contained in the out-of-sample sample. For this reason, we present the statistics of model performance for both samples. In case the relative advantage of one model is maintained across these two parts of samples, we can add confidence to our conclusion as to the competitive advantage of the winning model.

In table 4.1, we compare the in-sample forecasting performance of the Discrete-Event-Forecasting (DEF) model with that of the probit model. We choose different forecast leads, i.e., 0, 1, 3, 6, 9, 12. For each lead, we pick out the DEF model that has the optimal likelihood value. For each lead, we present two groups of statistics: (1) the goodness-of-fit measure which include the value of likelihood function and Quadratic Probability Score and (2) the turning point forecast performance measure which consists of four parts as

$$[N_{\text{PA}}(\text{Average Lead}), N_{\text{LA}}(\text{Average Lag}), N_{\text{FA}}, N_{\text{MA}}]$$

where PA=Prompt Alarm, LA=Late Alarm, FA=False Alarm, MA=Missed Alarm. An example, [1(-2),5(3.33),2,0], means that there is 1 prompt alarm which leads by 2 months, 5 late alarms, which on average lag by 3.33 months, 2 false alarm and 0 missed alarm. For each alarm threshold, we provide one set of such a measure. The threshold values we choose are 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.

First of all, both the Discrete-Event-Forecasting model and the probit model does better in long-horizon forecasts than in short-horizon forecasts. This observation confirms the previous findings that the financial variable does have a significant predictive power of business cycle regimes with a long lead.

Second, in short horizon forecasts, the Discrete-Event-Forecasting model does better if it puts more weight on the ARMA memory index of the level of Spread, i.e., $M_{y,1,t}$ as in (4.1). In long horizon forecasts, the Discrete-Event-Forecasting model does better if it puts more weight on the ARMA memory index of the change of Spread, i.e., $M_{y,2,t}$ as in (4.2). This phenomenon is consistent with the first observation. If we restrict the forecasting lead, k, to be short, the Discrete-Event-Forecasting model will automatically choose to put more weight and on the ARMA memory index of the level. In the mean time, it will select a large smoothing parameter for the ARMA memory index of the level. By doing so, the information of distant periods' term spreads' level can be used to predict the future regimes even if the lead is confined to be short.

Third, in terms of goodness-of-fit, the Discrete-Event-Forecasting model dominates the probit model when the forecasting lead is short or medium, i.e. shorter than 9. When the forecast lead is lengthened, the Discrete-Event-Forecasting model with a high or medium resolution level will do better than the probit model. This observation indicates that the flexibility of modelling does contribute a lot to the better performance of the Discrete-Event-Forecasting model.

In terms of peak turning point forecasting, the Discrete-Event-Forecasting model does better than the probit model, especially in the long horizon forecast. If we set the alarm threshold at 0.5 and set the horizon at 9 and 12, the Discrete-Event-Forecasting model will miss only 1 peaking points while the probit model will miss 3 turning points and 4 turning points, respectively.

The in-sample results in table 4.1 demonstrate that according to the in-sample performance, the Discrete-Event-Forecasting model yields a better model fit and better turning point forecasting performance. However, since we pick the best performer among all the Discrete-Event-Forecasting models to make comparison, the risk of overfitting is obvious. Through the out-of-sample study we try to examine whether this competitive advantage is robust.

In table 4.2, we compare the out-of-sample forecasting performance of the Discrete-Event-Forecasting model with the probit model. We choose the Discrete-Event-Forecasting models which yields the best in-sample performance to carry out the out-of-sample forecast.

In terms of goodness-of-fit, the Discrete-Event-Forecasting model dominates the probit model in all horizons at all the resolution levels. In terms of peak turning point forecasting, the Discrete-Event-Forecasting also does better than the Probit model in all horizons but horizon 1. For example, when we set the alarm threshold at 0.5, the probit model will miss the turning point in all horizons while Discrete-Event-Forecasting model will trigger the alarm promptly at all horizon except horizon 1.

This result suggests that Discrete-Event-Forecasting model can be used as an effective alternative method to the traditional probit model in predicting both business cycle regimes and business cycle turning points.

4.3.2 Plot

In figure 4.1, we draw plots of the fitted probability function $\Pr(S_{t+k} = 1|M_{y,t})$ and the fitted probabilities in both in-sample and out-of-sample forecasting. These plots for Discrete-Event-Forecasting model give evidence that the better performance of Discrete-Event-Forecasting model is not due to over-fitting. First, even the low resolution function estimator yields good probability fit for the recession regime switch. Second, for the out-of-sample forecasting, the Discrete-Event-Forecasting model shows greater sensitivity and triggers alert more promptly for turning point forecasts.

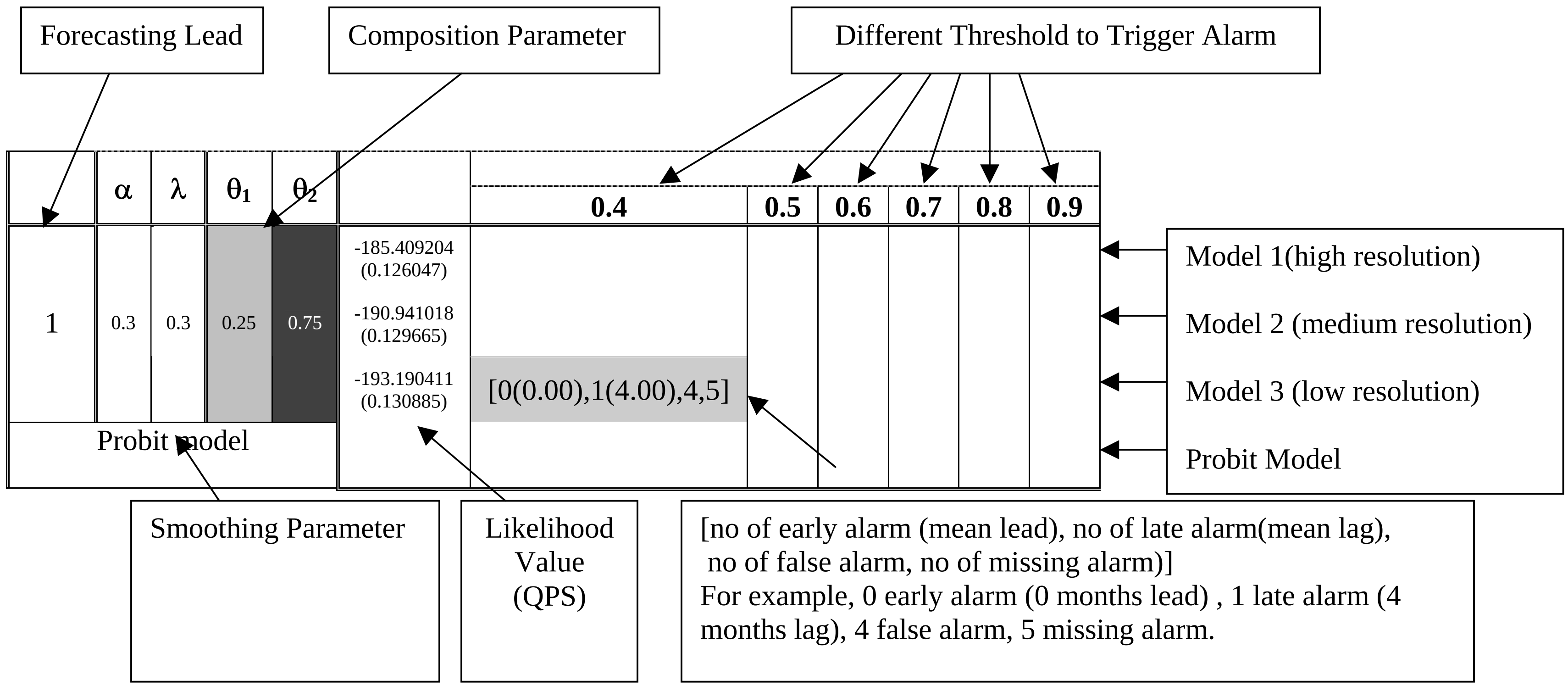
4.4 Conclusion

In business cycle regime forecast, the traditional probit model adopts a simple linear structure. It chooses the level of financial variables as information variables. In this section, we apply the Discrete-Event-Forecasting model to business cycle regime and turning point forecasts. Our study is motivated by the conjecture that it is not the level but the long-run trend of changes of financial variables that shed light on future business cycle regimes. In addition, since the probit model restricts the error terms of the latent variable to be normally distributed, deterioration of the forecasting performance is possible due to potential misspecification.

Through comparing the probit model with the Discrete-Event-Forecasting model in both in-sample and out-of-sample cases, we can see that the Discrete-Event-Forecasting model can do better than the traditional probit model. With respect to the overall model-fit, the difference between these two models will be obvious if we increase the flexibility of modeling. With respect to the turning point forecasting performance, the Discrete-Event-Forecasting model does much better. We can be released from the concern of the over-fitting problem since the better performance in turning point forecasts apply to functions of all resolution levels. This encouraging result shows the advantage of the new modeling approach in terms of the ability of detecting useful hints of the future rare event. It demonstrates the utility of Discrete-Event-Forecasting models in real-world practices.

Table 4.1 In-Sample Forecasting Performance Comparisons between the DEF model and the Probit Model

In table 4.1, we compare the DEF model with probit model on their in-sample forecasting performance of regime forecasts. For DEF model, we use different leads, e.g. 0, 1,3, 6, 9. For each lead, we pick out the model with the maximal likelihood function value. The dataset we apply is the interest rate term spread between 3 month Treasury bill and 10 years Treasury bill for 1959.8-1996.12.
Table Structure:



Lead	Para I		Para II		Performance						
	α	λ	θ_1	θ_2	Likelihood (MSE)	Threshold					
						0.4	0.5	0.6	0.7	0.8	0.9
1					-141.782915 (0.100892)	[3(-2.33),1(5.00),2,2]	[2(-3.00),1(6.00),1,3]	[1(-2.00),2(3.50),2,3]	[1(-2.00),2(6.00),0,3]	[0(0.00),1(0.00),0,5]	[0(0.00),1(3.00),0,5]
	0.9	0.9	0.75	0.25	-149.461711 (0.102923)	[3(-2.33),1(5.00),2,2]	[2(-3.00),1(6.00),1,3]	[1(-2.00),2(3.50),2,3]	[0(0.00),3(4.00),1,3]	[0(0.00),1(0.00),0,5]	[0(0.00),0(0.00),0,6]
					-156.507151 (0.108458)	[3(-2.33),1(5.00),2,2]	[0(0.00),1(0.00),0,5]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]
	Probit Model				-178.235437 (0.120819)	[0(0.00),1(3.00),2,5]	[0(0.00),0(0.00),1,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]
3					-133.563238 (0.094059)	[1(-7.00),3(0.33),2,2]	[1(-7.00),3(0.33),3,2]	[1(-6.00),1(0.00),1,4]	[1(-2.00),1(3.00),0,4]	[0(0.00),0(0.00),0,5]	[0(0.00),0(0.00),0,6]
	0.7	0.9	0.75	0.25	-136.535845 (0.095508)	[1(-7.00),3(0.33),3,2]	[1(-7.00),2(0.50),2,3]	[1(-6.00),1(0.00),1,4]	[1(-5.00),1(1.00),1,4]	[1(-2.00),1(7.00),1,4]	[0(0.00),0(0.00),0,6]
					-151.101546 (0.106128)	[1(-7.00),3(0.33),3,2]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]
	Probit Model				-161.087498 (0.110613)	[2(-1.00),1(1.00),2,3]	[0(0.00),3(0.33),2,3]	[0(0.00),2(1.50),2,4]	[0(0.00),2(3.00),1,4]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]
6					-128.506549 (0.090022)	[1(-5.00),3(1.33),4,2]	[1(-5.00),3(1.33),3,2]	[1(-4.00),3(3.33),2,2]	[0(0.00),2(3.50),2,4]	[0(0.00),1(2.00),1,5]	[0(0.00),1(3.00),1,5]
	0.7	0.9	0.75	0.25	-131.638742 (0.091976)	[1(-5.00),3(1.67),3,2]	[1(-4.00),3(2.67),2,2]	[1(-3.00),1(3.00),3,4]	[1(-2.00),1(4.00),1,4]	[0(0.00),2(3.50),2,4]	[0(0.00),0(0.00),0,6]
					-137.861904 (0.096629)	[1(-5.00),3(1.33),4,2]	[1(-4.00),2(3.50),3,3]	[0(0.00),2(3.00),1,4]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]	[0(0.00),0(0.00),0,6]
	Probit Model				-142.598147 (0.098059)	[1(-2.00),2(0.50),5,3]	[1(-2.00),2(2.00),3,3]	[1(-1.00),2(3.00),2,3]	[1(-1.00),1(5.00),1,4]	[0(0.00),0(0.00),1,5]	[0(0.00),0(0.00),0,6]

9	**0.9**	**0.9**	0.25	0.75	-131.894700 (0.090912)	[1(-3.00),5(1.60),8,0]	[0(0.00),5(2.40),5,1]	[0(0.00),4(2.75),1,2]	[0(0.00),3(2.67),2,3]	[0(0.00),1(0.00),0,5]	[0(0.00),1(0.00),0,5]
					-136.670054 (0.092295)	[1(-3.00),5(1.60),5,0]	[0(0.00),5(2.40),5,1]	[0(0.00),4(2.50),3,2]	[0(0.00),4(2.75),1,2]	[0(0.00),4(3.25),2,2]	[0(0.00),2(3.00),2,4]
					-140.409353 (0.094438)	[1(-3.00),5(1.60),5,0]	[0(0.00),5(2.20),5,1]	[0(0.00),4(2.25),3,2]	[0(0.00),4(3.00),2,2]	[0(0.00),3(4.00),1,3]	[0(0.00),1(6.00),2,5]
	Probit Model				-133.678970 (0.093155)	[1(-3.00),3(3.33),4,2]	[0(0.00),3(3.67),4,3]	[0(0.00),3(4.00),3,3]	[0(0.00),2(4.00),3,4]	[0(0.00),1(2.00),1,5]	[0(0.00),1(2.00),1,5]
12	**0.9**	**0.9**	0.25	0.75	-118.068435 (0.086502)	[2(-3.00),4(2.00),7,0]	[1(-2.00),5(2.00),5,0]	[0(0.00),1(0.00),0,5]	[0(0.00),1(0.00),0,5]	[0(0.00),1(0.00),0,5]	[0(0.00),1(0.00),0,5]
					-125.707404 (0.088861)	[1(-4.00),5(2.00),4,0]	[1(-1.00),4(3.25),4,1]	[0(0.00),3(4.00),3,3]	[0(0.00),3(4.33),1,3]	[0(0.00),3(4.67),2,3]	[0(0.00),1(0.00),0,5]
					-135.275474 (0.095217)	[1(-4.00),5(2.00),4,0]	[1(-1.00),4(5.25),4,1]	[0(0.00),2(5.50),4,4]	[0(0.00),2(6.00),1,4]	[0(0.00),2(6.50),1,4]	[0(0.00),1(9.00),2,5]
	Probit Model				-130.850679 (0.095040)	[0(0.00),4(4.75),4,2]	[0(0.00),2(6.00),4,4]	[0(0.00),2(6.50),2,4]	[0(0.00),1(5.00),3,5]	[0(0.00),1(5.00),1,5]	[0(0.00),1(5.00),1,5]

Table 4.2 Out-of-Sample Forecasting Performance Comparisons between the DEF model and the Probit Model

In table 4.2, we compare the out- of- sample forecasting of DEF model with probit model. The structure of the table is same as that of table 4.1. The dataset we apply is the interest rate term spread between 3 month Treasury bill and 10 years Treasury bill for 1997.01-2001.12.

lead	Model	Forecasting Performance						
		Likelihood (MSE)	Threshold					
			0.4	0.5	0.6	0.7	0.8	0.9
1	DEF	26.784288 (0.135945)	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		26.140758 (0.133296)	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		26.591100 (0.135635)	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
	Probit	29.895752 (0.143158)	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
3	DEF	25.641708 (0.132571)	***[1(-4.00),0(0.00),0,0]***	***[1(-3.00),0(0.00),0,0]***	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		26.105717 (0.135790)	***[1(-3.00),0(0.00),0,0]***	***[1(-3.00),0(0.00),0,0]***	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		27.689880 (0.138181)	***[1(-3.00),0(0.00),0,0]***	***[0(0.00),0(0.00),0,1]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
	Probit	29.905626 (0.144667)	[0(0.00),0(0.00),1,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
6	DEF	17.765227 (0.079307)	***[1(-1.00),0(0.00),0,0]***	***[1(-1.00),0(0.00),0,0]***	***[0(0.00),1(0.00),0,0]***	***[0(0.00),0(0.00),0,1]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		17.808814 (0.079335)	***[1(-1.00),0(0.00),0,0]***	***[0(0.00),1(0.00),0,0]***	***[0(0.00),1(1.00),0,0]***	***[0(0.00),1(2.00),0,0]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		19.593348 (0.091829)	***[1(-1.00),0(0.00),0,0]***	***[0(0.00),1(0.00),0,0]***	***[0(0.00),1(2.00),0,0]***	***[0(0.00),0(0.00),0,1]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
	Probit	23.908606 (0.111860)	[0(0.00),1(0.00),1,0]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
9	DEF	14.603846 (0.070332)	[0(0.00),1(0.00),2,0]	***[0(0.00),1(1.00),0,0]***	***[0(0.00),1(5.00),0,0]***	***[0(0.00),0(0.00),0,1]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		13.918791 (0.065761)	[0(0.00),1(1.00),0,0]	***[0(0.00),1(1.00),0,0]***	***[0(0.00),1(4.00),0,0]***	***[0(0.00),1(5.00),0,0]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		13.346295 (0.063410)	[0(0.00),1(1.00),0,0]	***[0(0.00),1(1.00),0,0]***	***[0(0.00),1(3.00),0,0]***	***[0(0.00),1(5.00),0,0]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]

	Probit	17.015548 (0.079311)	[0(0.00),1(1.00),0,0]	***[0(0.00),0(0.00),0,1]***	***[0(0.00),0(0.00),0,1]***	***[0(0.00),0(0.00),0,1]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
12	DEF	15.473879 (0.077876)	[0(0.00),1(3.00),2,0]	***[0(0.00),1(4.00),0,0]***	***[0(0.00),0(0.00),0,1]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		15.158209 (0.072550)	[0(0.00),1(4.00),0,0]	***[0(0.00),1(4.00),0,0]***	***[0(0.00),1(7.00),0,0]***	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
		14.085785 (0.068270)	[0(0.00),1(4.00),1,0]	***[0(0.00),1(4.00),0,0]***	***[0(0.00),1(6.00),0,0]***	[0(0.00),1(8.00),0,0]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]
	Probit	17.288476 (0.082387)	[0(0.00),1(4.00),0,0]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]	[0(0.00),0(0.00),0,1]

Figure 4.1 Plots of Estimation and Forecasting Results of the Discrete-Event-Forecasting Model for Business Cycle Regime Forecasts, Forecasting Lead=1

In figure 4.1-4.4, we plot (1) the fitted probability for in-sample data (2) the estimated conditional probability and (3) the probability forecasted for out-of-sample. There are 4 sets of plots in total. And for each set, we have 9 plot with 3 for each model.

0_0_FYGM3_0_FYGT10(1959.08-1996.12, in sample, 1997.01-2001.12 out of sample)

The Fitted Probability For In-Sample | The Conditional Probability function | Out-of-sample forecasting Result

Model 1
(High Resolution)

Model 2
(Med Resolution)

Model 3
(Low Resolution)

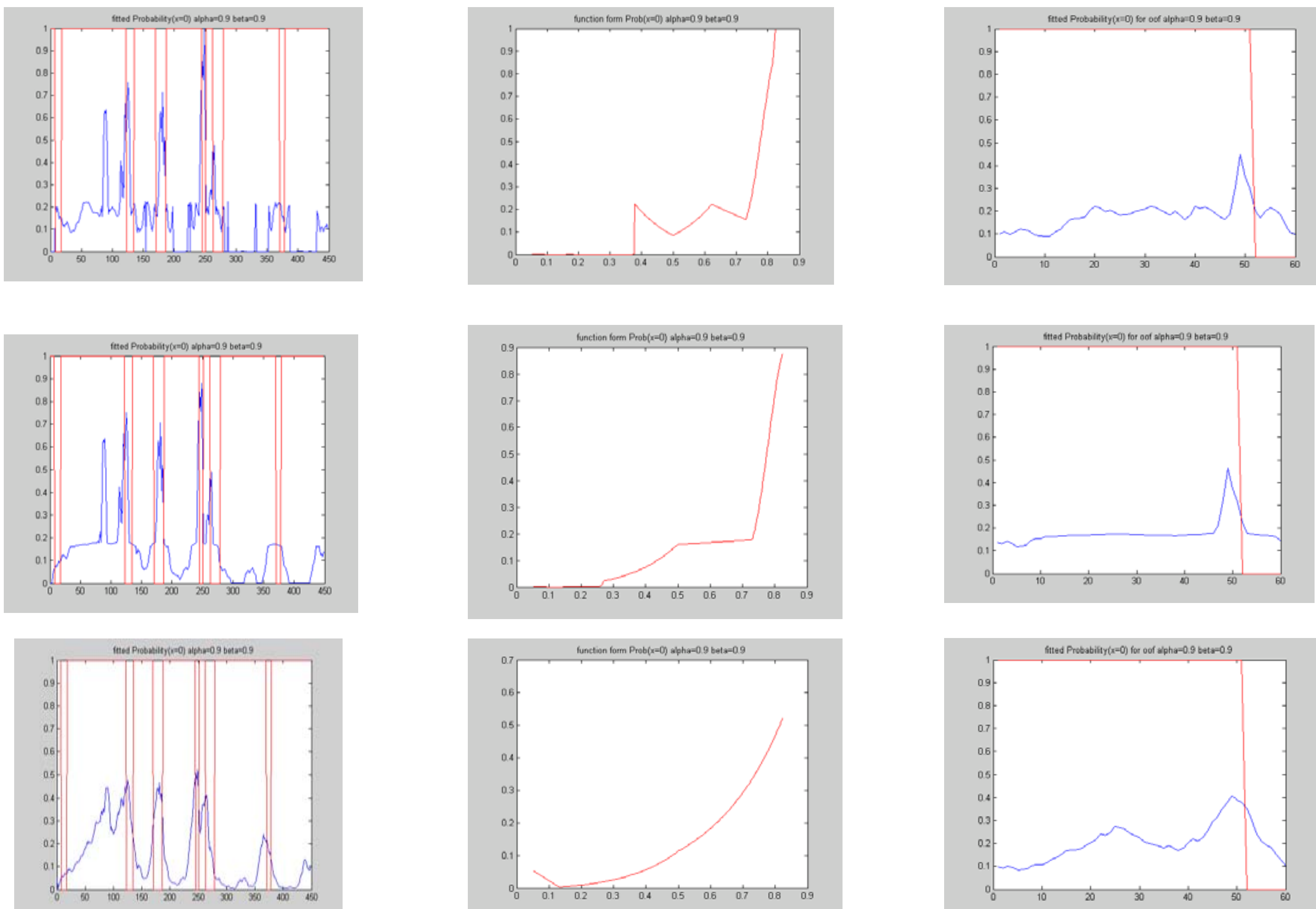


Figure 4.2 Plots of Estimation and Forecasting Results of the Discrete-Event-Forecasting Model for Business Cycle Regime Forecasts, Forecasting Lead=3

0_0_FYGM3_0_FYGT10(1959.08-1996.12, in sample, 1997.01-2001.12 out of sample)

The Fitted Probability For In-Sample | The Conditional Probability function | Out-of-sample forecasting Result

Model 1
(High Resolution)

Model 2
(Med Resolution)

Model 3
(Low Resolution)

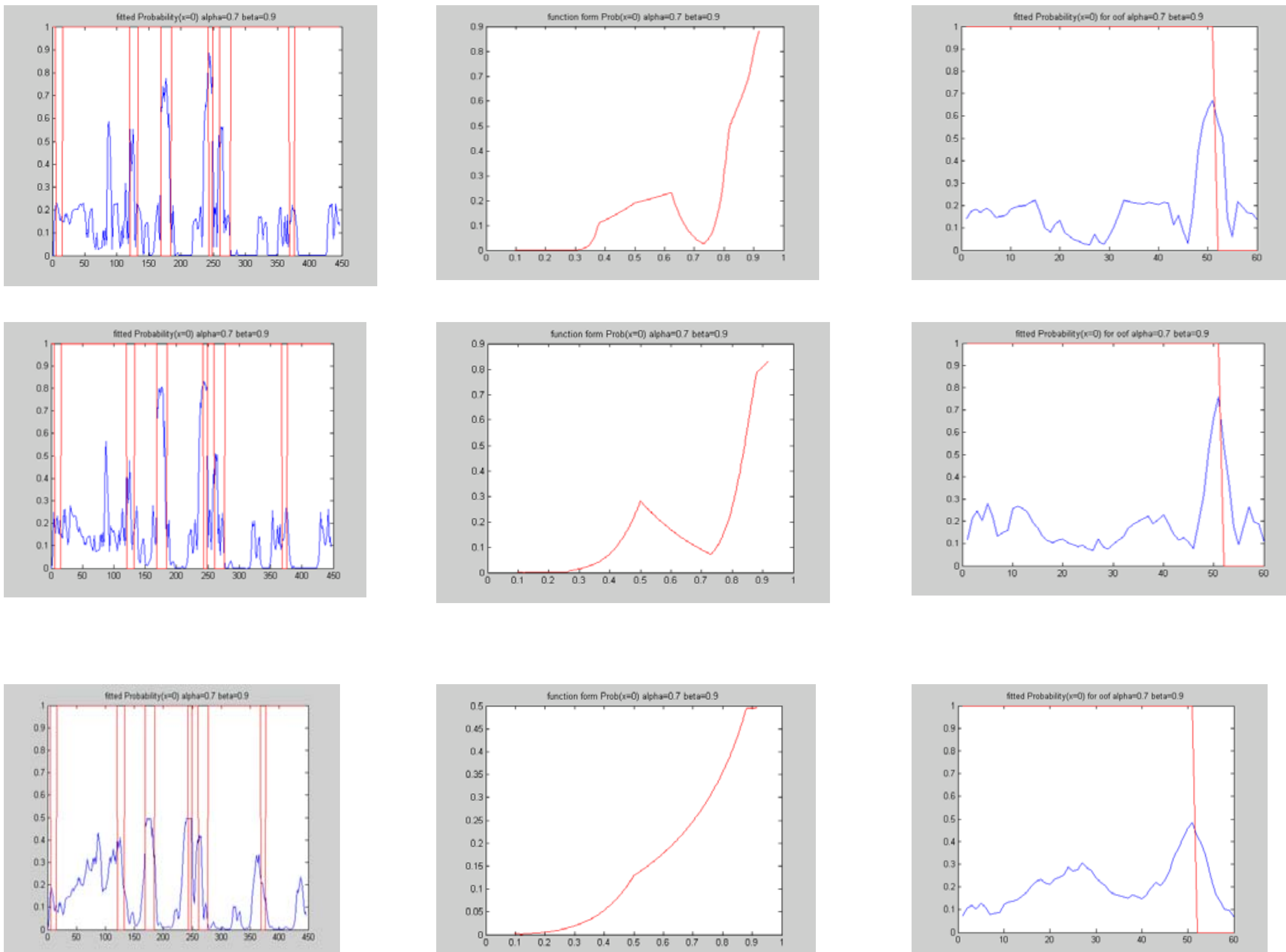


Figure 4.3 Plots of Estimation and Forecasting Results of the Discrete-Event-Forecasting Model for Business Cycle Regime Forecasts, Forecasting Lead=6

0_0_FYGM3_0_FYGT10(1959.08-1996.12, in sample, 1997.01-2001.12 out of sample)

The Fitted Probability For In-Sample | The Conditional Probability function | Out-of-sample forecasting Result

Model 1
(High Resolution)

Model 2
(Med Resolution)

Model 3
(Low Resolution)

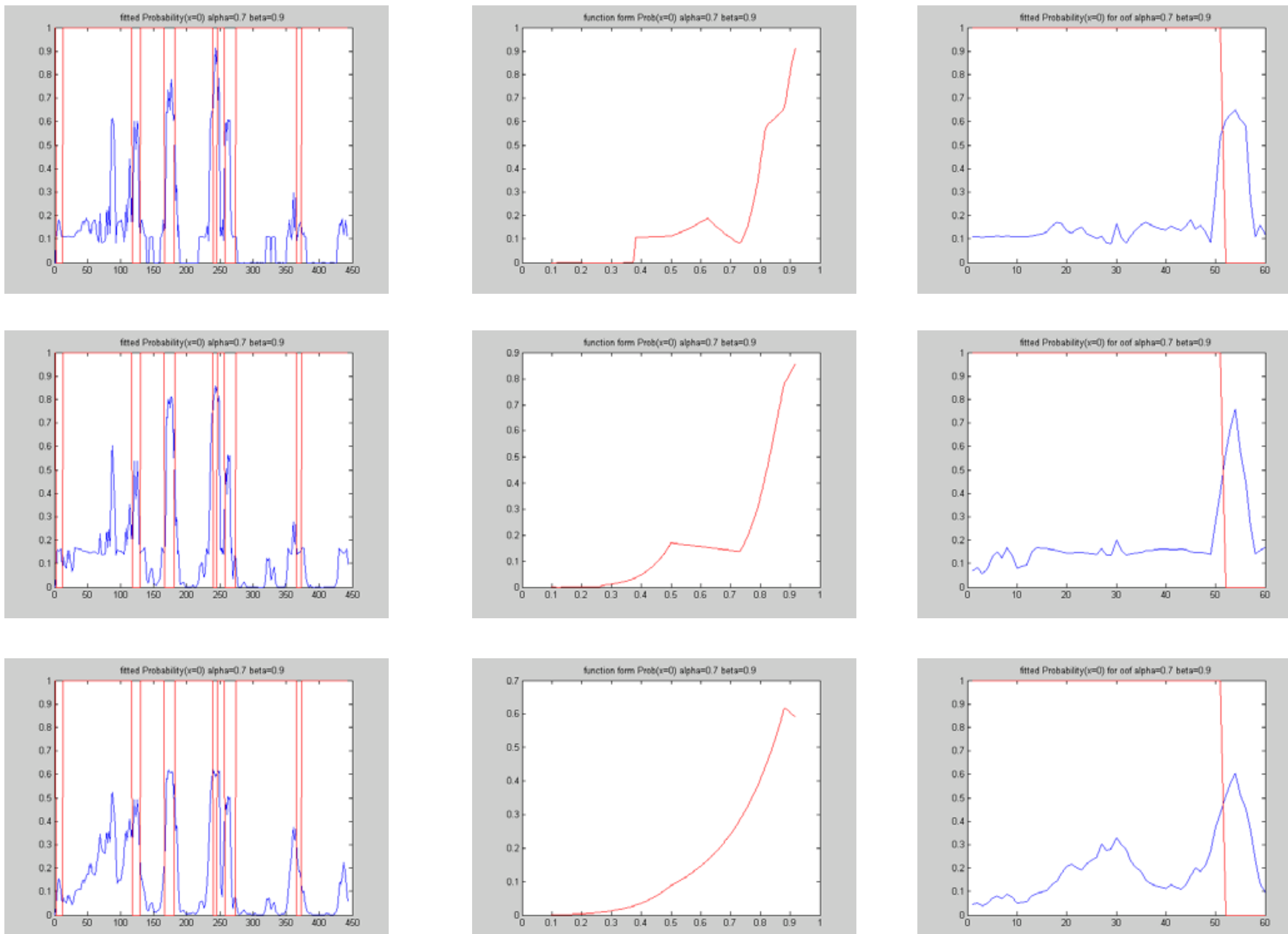


Figure 4.4 Plots of Estimation and Forecasting Results of the Discrete-Event-Forecasting Model for Business Cycle Regime Forecasts, Forecasting Lead=9

0_0_FYGM3_0_FYGT10(1959.08-1996.12, in sample, 1997.01-2001.12 out of sample)

The Fitted Probability For In-Sample | The Conditional Probability function | Out-of-sample forecasting Result

Model 1
(High Resolution)

Model 2
(Med Resolution)

Model 3
(Low Resolution)

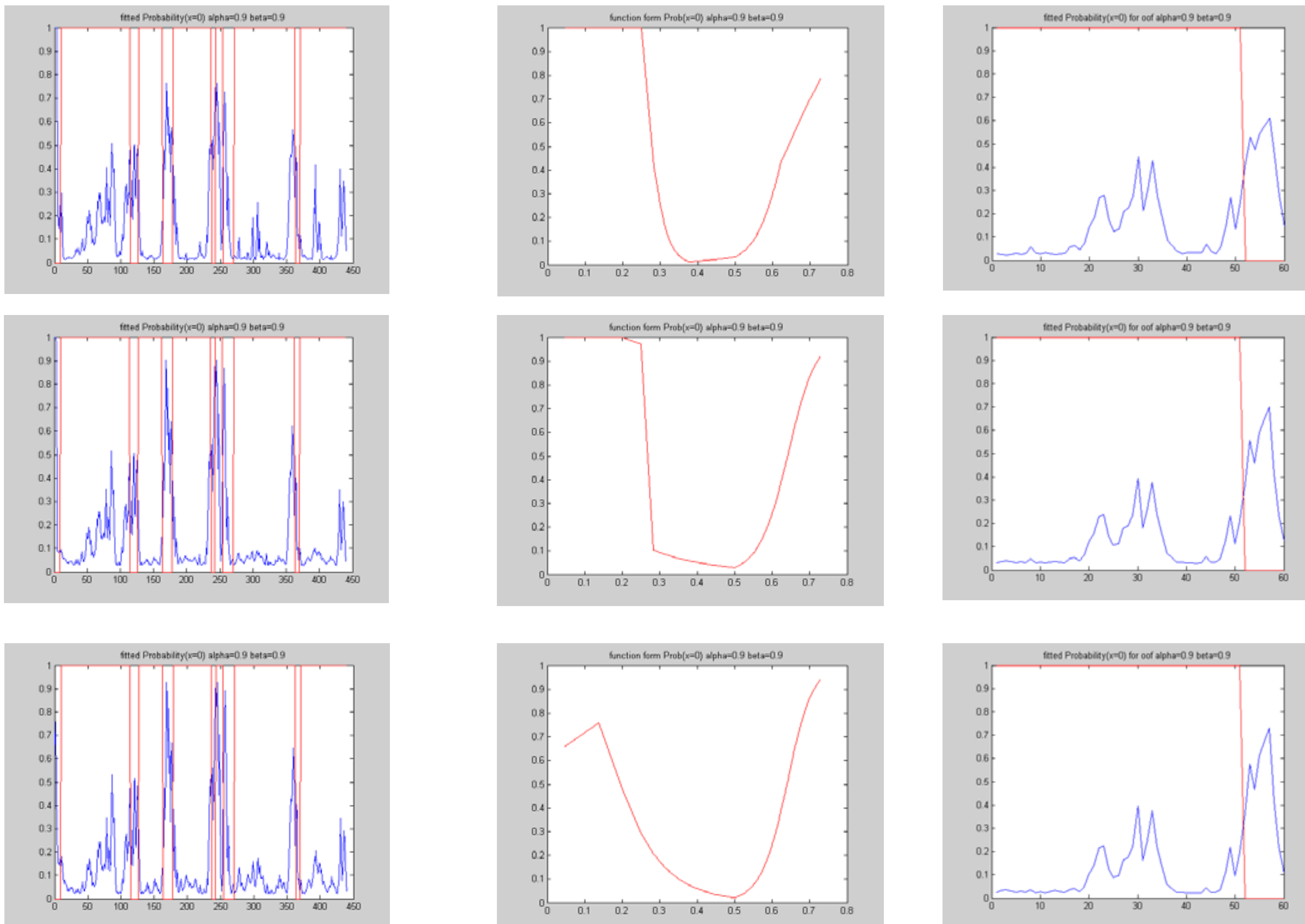


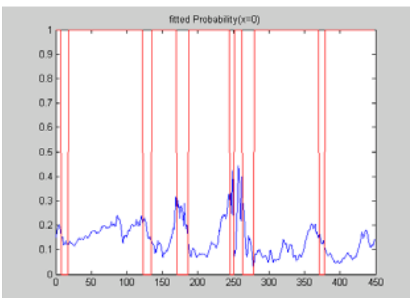
Figure 4.5 Plots of Estimation and Forecasting Results of the Probit Model for Business Cycle Regime Forecasts

In figure 4.5, we plot (1) the fitted probability for in-sample data (2) the probability forecasted for out-of-sample. There are 4 sets of plots in total. And for each set, we have 2 plots.

0_0_FYGM3_0_FYGT10(1959.08-1996.12, in sample, 1997.01-2001.12 out of sample)

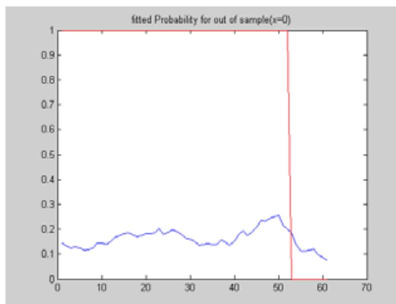
The In-Sample Fitted Probability

Lead=1



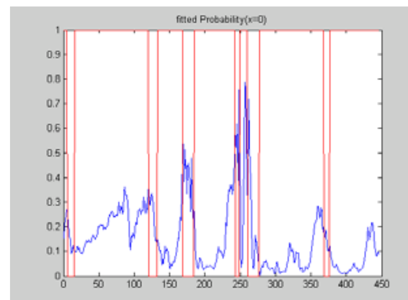
Out-of-sample forecasting

Lead=1



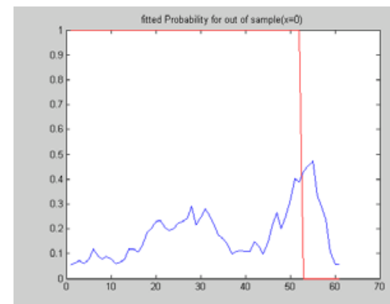
The In-Sample Fitted Probability

Lead=3



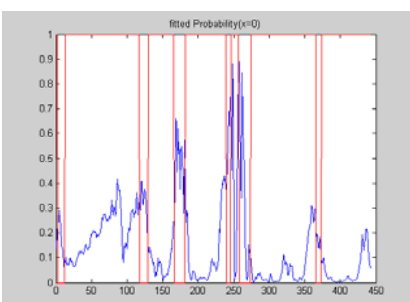
Out-of-sample Forecasting

Lead=3



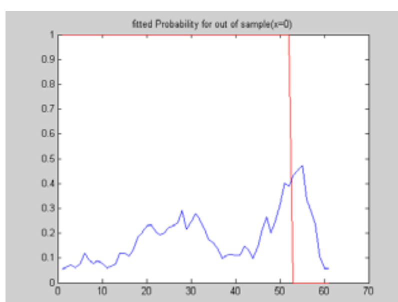
The In-Sample Fitted Probability Graph

Lead=6



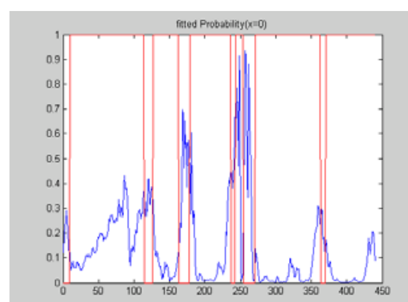
Out-of-sample forecasting

Lead=6



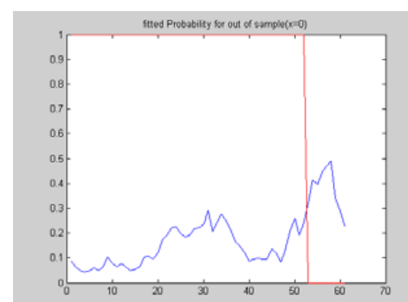
The In Sample Fitted Probability

Lead=9



Out-of-sample forecasting

Lead=9



Chapter 5

APPLICATION OF DISCRETE-EVENT-FORECASTING MODEL IN STOCK RETURN FORECAST

In the study of stock return forecast, there are two main branches. One focuses on the conditional mean forecasting (see Cootner 1964, Lo 1997) and the other on the conditional variance forecasting (see Bollerslev et al 1992 and Anderson et al 2003)[1]. The probabilistic forecasting of individual stock return events such as a stock outperforming a benchmark index on monthly returns or a stock return's performance being ranked the lowest during the period t, however, has received much less attention, even if it is of great importance to both investing practices and academic researches.

First, accurate forecasts of discrete stock return events could benefit stock portfolio management greatly. The probabilistic forecasting of discrete stock return events will provide more accurate information of the stochastic properties of stock returns than the moment forecasting does, if the distribution function of return variables departs markedly from a normal one. A

[1] The literature to which we refer mainly focus on the predictability of individual stock returns when conditioning information only includes price information. Studies in this aspect directly attack market efficiency. There are abundant evidences of predictability of portfolios' returns when we can condition on historic information of dividend yields, interest term spreads and etc. (see Campbell 1987, Fama & French 1988, Ferson & Campbell 1999). Such predictability can be rationalized via the intertemporal capital asset pricing model (ICAPM) developed in Merton (1973).

forecasting model that directly tackles the problem of probabilistic forecasting of stock return events is appealing to investors due to its robustness against serious model misspecification on the conditional moment functions of the underlying data-generating-process of stock returns.

Second, for academic researches, studies in the predictability of discrete stock return events can help reveal the source of the predictability of certain moment functions and solve long lasting puzzles in the asset pricing literature. In previous studies, price reversal and price continuation have been both discovered in the stock market (see De Bondt and Thaler 1985,1987, Lehmann 1990, Jagadeesh 1990, Jagadeesh and Titman 1993). These findings indicate a noticeable relationship between stock price histories and expected stock returns.[2] Due to these empirical results, behavioral finance, the area that studies market inefficiency, has grown into a promising research field in the last decade (see Shleifer 2000 and Shefrin 2000). However, the predictability of the conditional mean may not necessarily imply pricing inefficiency if it is accompanied by the predictability of higher-order moments which are risk factors (see Ferson & Harvey 1991,1999, Harvey and Siddique 2000). Studies in stock return events forecast will provide an effective way to exploit the predictability of various moment functions, thus improving the understanding of the market efficiency.[3]

One main issue in the study of probabilistic forecasting is modeling consistency. Predictability of stock return events reflects the dependence of a stock return's distribution function on its entire past. But we do not have good prior knowledge on the lag structure of such dependence.[4] Nor do we have established tests for the lag selection. Therefore, any lag truncation could cause model misspecification. Moreover, various dynamic features of different moment functions of a stock return will have a joint impact on the predictability of specific stock return events, thus increasing the complexity of such dependence. Model misspecification could well happen with parametric modeling methods that set restrictions on the functional form of the conditional probability functions. Probabilistic forecasting of discrete stock return events will

[2] For example, the momentum effect in Jagadeesh and Titman (1993) indicates that a strong positive correlation exists between historic price information and future stock returns. The fads phenomenon in Lehmann (1990) shows that there is a significant negative serial auto-correlation for weekly stock return series.

[3] McQueen and Thorley (1991) use a Markov chain model to test the random walk hypothesis of stock prices. The model is used to estimate the transitional probability between two types of events, a high return event and a low return event. Their study shows that the movement of annual real returns exhibits a significant deviation from the random walk hypothesis.

[4] As shown in Jagadeesh(1990), for monthly stock return series, even the 12th lag can have a significant impact.

generally yield unsatisfactory performance due to inconsistent modeling.

In this section, we attempt to apply discrete events forecasting model to conduct probabilistic forecast of interesting discrete stock return events. We try to examine the possibility of improving the forecasts of stock return events and gaining a more accurate measurement of the uncertainty embedded in stock return series.

5.1 Research Design

5.1.1 Data

The database that we use is the monthly stock price database provided by the Center for Research in Security Prices(CRSP) with the time range from 12/31/1925 to 12/31/2002. We choose monthly price data of stocks that are listed in NYSE+AMEX only. In addition, any stock that is admissible into our samples must

(1) be an ordinary common stock which has not been or need not be further defined by the database from the Centre of Research in Stock Price.[5]

(2) have a price above \$5 at the period that we pick for data sampling.[6]

(3) have no less than 18 observations before the period that we pick for data sampling.

In addition to monthly stock price data, we use the monthly price of S&P 500 composite index, ranging from 12/31/1925 to 12/31/2002, to construct our benchmark index return. The data is also provided by CRSP.

5.1.2 Forecasting Events

At each time period t, we forecast the event that an individual stock's return in the next period will exceed that of the benchmark index, S&P 500 composite index. Formally, let $\{R_{i,t}\}_{t=1}^{\infty}$ be the series of the stock i's holding return (excluding dividend) during the periods t. let $\{R_{0,t}\}_{t=1}^{\infty}$ be the holding return for the S&P 500 composite index (excluding dividend) during the period t. We define the event variable $E_{i,t}$ as

[5] In database from CSRP, these stocks have share code 10 or 11. It excludes ADRs,SBIs,Units,certificates, Foreign Stocks, American Trust Components, close-end funds and REIT's.

[6] Stocks whose prices are lower than \$5 are known as Penny Stocks. Dropping them will prevent the outcome of empirical study from being overly affected by outliers.

$$E_{i,t} \quad \begin{cases} = 1 & \text{if } R_{i,t} > R_{0,t} \\ = 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Our aim is to estimate

$$\Pr(E_{i,t} = 1 | \mathcal{F}_{t-1}) \tag{5.2}$$

in order to do probabilistic forecasting of events $R_{i,t} > R_{0,t}$ conditional on information up to time period $t-1$, $\mathcal{F}_{t-1}$.

We choose this event to forecast for the following reasons. First, knowing how likely a stock's return is to beat an benchmark index in the near future is important for fund portfolio managers. One crucial indicator of fund managers' capability is whether they could manage their funds to yield returns that consistently exceed that of a benchmark index. Second, many studies in stock return predictability attempt to find a trading strategy that has the ability to distinguish stocks that will outperform benchmark indexes from those that will underperform. In our study, we define a target event such as (5.1) so that we can directly investigate the feasibility of implementing such a strategy.

5.1.3 Data Clipping

As discussed in section 1, in order to implement discrete event forecasting modelling, we need to discretize the original conditioning random variables in the forecasting model.

First, we pick a universe of stocks to formulate a stock base S, i.e., domestic stocks listed in NYSE and AMEX during 12/31/1925-12/31/2002. For each month during 12/31/1925-12/31/2002, we rank all of the stocks available during that month according to $R_{i,t}$. A new discrete-valued return series $\{R^r_{i,t}\}_{t=1}^{T_i}$ is defined according to the monthly return rank group to which $R_{i,t}$ belongs. More precisely,

$$R^r_{i,t} = \sum_{j=1}^{10} jI(R_{i,t} \in [d_{j-1,t}, d_{j,t})), i \in S \tag{5.3}$$

where $\cup_{j=1}^{10}[d_{j-1,t}, d_{j,t}) = \mathbb{R}$, and $d_{j-1,t} < d_{j,t}$, $j = 1, ..., 10$.

Through this data transformation procedure, the history of stock returns becomes the history of a stock return's rankings. We consider the ranking information as informative about the future stock returns due to the findings in Jagadeesh and Titman (1993).

5.1.4 Modelling Wavelet-Based Conditional Probability Functions

Using a discrete-valued return series $\{R^r_{i,t}\}_{t=1}^{T_i}$, we construct the ARMA memory index $M_t(\alpha)$ as follows

$$M_{i,t}(\alpha) = \sum_{j=1}^{\infty} \alpha^{j-1} R^r_{i,t-j} \text{ and } \alpha \in (0,1) \tag{5.4}$$

Then we transform the conditional probability function of the event (5.1) from (5.2) into

$$\begin{aligned} \Pr(E_{i,t} &= 1|R^r_{i,t-1}, R^r_{i,t-2}, ...) \\ &= \Pr(E_{i,t} = 1|M_{i,t}(\alpha)) \\ &= F\left[M_{i,t}(\alpha)|\alpha\right] \end{aligned} \tag{5.5}$$

where $F(x)$ takes the form of a logit function

$$F(x) = \frac{1}{1 + \exp(-g(h(x)))} \tag{5.6}$$

with

$$h(x) = \ln(\frac{x}{1-x}), x \in (0,1).$$

In (5.6), $g(x)$ is a real-valued function with the domain $\mathbb{R}$, which has a wavelet representation such as

$$g(x) = \sum_{k=-k_0}^{k_0} \beta_{J_0,k}\varphi_{J_0,k}(x) + \sum_{j=J_0}^{J_1} \sum_{k=-k_{1,j}}^{k_{2,j}} \beta_{j,k}\psi_{j,k}(x) \tag{5.7}$$

where $\varphi_{j,k}(x)$ is the scaling function and $\psi_{j,k}(x)$ is the mother wavelet function of a particular wavelet basis. In this paper, we use Franklin Wavelets, also known as Spline Wavelet with polynomial degree 1, as the basis for the functional space (see the appendix for detail on wavelets construction). We choose three combinations of wavelet components to construct the function in (5.7). The detail is the same as in the chapter 3.

5.1.5 Sampling Data

We use three different methods to draw learning samples for model estimation and do out-of-sample forecasting. We do random sampling for all of the three methods in order to avoid the problem of data mining. Unlike previous studies which need to use new data to verify their findings, it is easy to check the robustness of our results by just doing another round of sampling.

1) Time Series Method—Among all of the admissible stocks,[7] we randomly pick 200 stocks. For each stock, we use the first two thirds of its total observations to estimate the conditional probability function (5.5). The remaining observations are used for out-of-sample forecasting.

2) Time Varying Cross Sectional Method—We select 200 samples of data in the following way. To construct each sample d, we randomly pick a month t_d from 01/1946 to 12/2002. We use historic return data on all of the admissible stocks that have valid return data up to the month $t_d + 1$ to formulate the dth sample. That is, we use all of the historic monthly return data up to the month t_d to construct the ARMA memory index $M_{i,t_d}(\alpha)$ and we use the return data of the month $t_d + 1$ to formulate the event variable, E_{i,t_d+1}. Then we use this sample data to estimate the conditional probability function (5.5). We use the same way to formulate the first round and second round dth out-of-sample forecasting sample by using all of the historic monthly return data up to the month $t_d + 2$ and $t_d + 3$ respectively.

3) Time Invariant Cross Sectional Method—Unlike the time varying cross sectional method, in this sampling method, we do not require that the data observations of the same sample be from the same time period. The drawing procedure is as follows. For the dth sample, we first randomly draw a month $t_{i,d}$ during 01/1946-12/2002 N_d times. Therefore, $i = 1, ..., N_d$ for the dth sample.For each picked month $t_{i,d}$, we randomly choose a stock $j_{i,d}$ that has valid price data during the month $t_{i,d}$ and $t_{i,d} + 1$. Then we formulate the dth learning sample in the same manner as in the time varying cross sectional method. We do this repeatedly for the other samples until we get 200 samples in total. We control the size of each sample in the following way. For the dth sample in the time invariant cross sectional method, we draw the

[7] Apart from those general selection rules in section 2.1, we require that the number of observations for each stock picked is no less than 500 during 12/31/1925-12/31/2002. There are 571 stocks in the stock base satisifying this requirement.

same number of observations as we do for the dth sample in the time varying cross sectional method. In addition, the dth sample in the time invariant cross sectional method shares the same out-of-sample forecasting sample with the dth sample in the time varying cross sectional method.

For the time invariant cross sectional sampling method, we can easily expand the learning samples for each out-of-sample forecasting sample. In case we want a larger learning sample, we just need to combine several learning samples together to do model estimation. However, in order to avoid spurious forecasting results, we put 30 as the limit of expansion.[8]

We draw sample data using these three different methods for the purpose of making comparisons and gaining a better understanding of the characteristics of the predictability of discrete events.

If data samples are drawn using the time series method, the data observations within the same sample belong to the same stock. Predictability of stock return events, if found, is considered a stationary time series property of the underlying data generating process of the specific return series involved. Therefore, positive findings using this method give the strongest support to the claim that the probabilities of stock return events depend on a stock's historic price information.

If data samples are drawn using the cross sectional methods, the data observations within each sample belong to different stocks. Predictability of stock return events, if found, can be due to either the time series properties of each individual stock or the cross sectional difference in expected returns among stocks.

The time varying cross sectional method uses local information to construct a conditional probability function while the time invariant cross sectional method uses global information to estimate the forecasting model. If the strength of the predictability is locally persistent, models that are estimated using the time varying method will yield better forecasting performance. If the strength and content of the predictability changes dramatically over time, we have to turn to global information to capture such predictability. Under that circumstance, models that are estimated using the time invariant method tend to provide a better forecast.

[8] Since the data is drawn globally, with 1 samples for model learning, on average, we will have 2 pairs of observations shared by the out-of-sample forecasting sample and in-sample learning. The regular size of a out-of-sampl forecasting sample is around 1000-1200.

5.1.6 Out-of-Sample Forecasting

First, for the cross-sectional sampling methods, due to the easy control on the data availability, we design a two-round out-of-sample forecasting scheme. After we estimate the 30 conditional probability function models, we do first round out-of-sample forecasting. Based on the first round out-of-sample forecasting performance, we pick a sampling method between the time varying cross sectional method and the time invariant cross sectional method. In the meantime, we select the optimal forecasting model from all of the 30 models. Then we do the second round out–of-sample forecasting to verify the consistency between the two rounds of out-of-sample forecasting. For the time series method, due to the limitation in the data size, we only carry out one round out-of-sample evaluation.

Second, we choose different regions in the functional domain of the conditional probability function (5.5) to do out-of-sample forecasting. That is, we want to check whether the discrete events forecasting models will have different forecasting performance against the reference model when the ARMA memory index falls into a different regions of the functional domain. The strength and content of predictability of stock return events may not be uniform across regions. As a result, the relative model performance of forecasting models may not be uniform across regions, either. To find the optimal way of using our forecasting models, we need check the variation of model performance across different regions of the functional domain.

The choice of the regions is not arbitrary. Intuitively, the more impressive a stock's historic return is, the more likely it is that the past return contains some useful information about the future return event. Therefore, we choose different end regions of the functional domain to check the existence of predictability. For example, after we estimate the functional form and do out-of-sample forecasting, we choose the a% end regions to do forecasting. That is, we choose stocks whose ARMA memory index's value falls into the bottom a% group or the top a% group among all of the admissible stocks in each forecasting period.

5.1.7 Evaluation Method

We use a standard forecasting evaluation approach which consists of three part, model learning, model selection and out-of-sample evaluation.

First, we estimate the conditional probability function by adopting a standard Maximum

Likelihood Estimation method. After we obtain the functional estimator, we then do the first round out-of-sample forecasting. The performance measure used in our forecast evaluation is Quadratic Probabilities Score (QPS), a standard measure for event forecasting evaluation. It can be expressed as

$$QPS_d = \frac{\sum_{i=1}^{I_d}(E_{i,t_d} - \Pr(E_{i,t_d} = 1|\mathcal{F}_{t-1}))^2}{I_d} \tag{5.8}$$

where I_d is the total number of stocks in the dth sample, $\Pr(E_{i,t_d} = 1|\mathcal{F}_{t-1})$ is forecasted value of conditional probabilities for event E_{i,t_d}.

Since this is a small sample case and we do not have good knowledge of the optimal expansion size for wavelet components, we use different combinations between wavelets components and ARMA memory index coefficients, α in $M_{i,t}(\alpha)$, to construct different models. We use 10 values for α, i.e., $0.1, ..., 0.9, 0.95.$,and 3 choices of wavelets components (see section 1.2). Consequently, we have 30 sets of out-of-sample forecasting results from discrete events forecasting models.

In the meantime, we set up different reference models corresponding to the three different sampling methods to check whether the conditional probability model using discrete events forecasting modeling approach could outperform a naive probability model which adopts a random walk hypothesis.

• For the time series method, we use a stock-specific constant probability to construct the stock specific random walk model as the reference model. For each stock i in the sample d, we use historic data observations of this stock up to the forecasting period to compute the fraction of observations for which $R_{i,t} > R_{0,t}, t = 1, ..., t_d$. We use this fraction as the estimator of the unconditional probability, $\Pr_i(R_{i,t} > R_{0,t})$,for each stock i for sample d.

• For the time varying cross sectional method, we use a period-specific constant probability to construct the period specific random walk model as the reference model. For each sample d , we use all of the stock data observations of the month t_d to compute the fraction of stocks for which $R_{i,t_d} > R_{0,t_d}$, $i = 1, ..., I_d$, where I_d is the total number of stocks in the dth sample. We use this fraction as the estimator of $\Pr_{t_d}(R_{i,t} > R_{0,t})$ for each sample d.

• For the time invariant cross sectional method, we use a time-invariant constant probability to construct the time invariant random walk model. For each sample d, the reference model is estimated using the same procedure as in the time varying cross sectional case. However, since samples are drawn globally, the probability estimators contain few time-specific factors.

Second, we do the first round out-of-sample evaluation. This step allows us to examine whether our new models have a potential to beat the reference model in out-of-sample forecasting. It also gives us the chance of selecting the optimal model among all of the competing models. In the meantime, we apply the Superior Predictive Ability test developed in Hansen (2001) to examine whether a reference model has better expected forecasting performance than all of the competing models. This test will be particularly useful if we cannot apply second-round out-of-sample forecast due to the limited sample size.

Formally, we define $T^*_{\max}$ as

$$T^*_{\max} = \max_m \frac{E\left(\nabla MS_m\right)}{\sigma_{mm}}$$

where $\nabla MS_{i,m} = QPS_{i,m} - QPS_{i,0}$ is the performance difference between the mth competing model and the reference model and σ_{mm} is the standard deviation of the sampling error for the mth model. The null hypothesis of SPA test is

$$H_0 : T^*_{\max} \leq 0 \tag{5.9}$$

The main purpose of this test is to minimize the risk of spurious good performance of alternative models (see Sullivan et al 2001 for an alternative approach). It adopts the standard stationary bootstrap technique developed in Politis & Romano(1994) to obtain the p-value of the null hypothesis. Since we have constructed a system of forecasting models using the discrete events forecasting modelling approach, the Superior Predictive Ability test is a natural choice. We refer interested readers to Hansen (2001) for more detail and in the appendix we give an example of the statistics construction for one of our sampling methods.

For the different sampling methods, we will apply the Superior Predictive Ability test in a different way. For the two cross sectional methods, we compare forecasting models based on the overall performance on all of the 200 samples. This is because the forecasting model is not stock-specific and the overall performance on all of the 200 samples gives an appropriate indicator of long run performance of each forecasting model. For the time series method, however, we compare forecasting models sample by sample because the optimal forecasting model is obviously stock specific.

Given that a positive testing result is obtained in the first round evaluation, we will continue with the second round out-of-sample evaluation. As mentioned above, due to the limitation in the sample size, we will not do the second round out-of-sample evaluation for the time series sampling method.

Third, for the cross sectional sampling methods, based on the evaluation results in the first round out-of-sample forecasting, we choose the optimal forecasting model and do the second round out-of-sample forecasting. We use the outcome of the second round out-of-sample forecasting to examine the informativeness of the evaluation results for the first round out-of-sample forecasting.

5.2 Result

5.2.1 Time Series Method

In table 5.1, we compare the one-month-ahead out-of-sample forecasting performance between the stock-specific random walk model and the 30 discrete events forecasting models.

As mentioned above, we conduct the Superior Predictive Ability test stock by stock. That is, for each stock, we test the null hypothesis that the reference model performs better than all of the competing models. In table 5.1, we provide information on the frequency that we reject the null hypothesis out of 200 samples. We also provide the performance statistics to show the average improvement made by competing models when the null hypothesis is rejected. In the same table, we provide the same statistics for each sub sample. The sub samples differ from each other in the basic attributes of the stocks, i.e., market size, beta, and standard deviations of stock returns.

The results for the whole sample analysis show that the competing models outperform the reference model significantly for 40% of all of the sampled stocks. The average improvement in Quadratic Probability Score is 0.00378 with standard error 0.00014. It should be noted that due to the requirement of the minimal history length, there might be a survivorship bias in the above results. Most stocks with a long history have at least above-average ranks in the market size, beta and standard deviation of returns. The average ranking in size, beta, and standard deviation of the stocks we pick in our samples are 7.86, 5.87, 7.26, out of 10 respectively. Consequently, stocks with a long history, on average, will be less risky. Trading on such stocks is usually more liquid and pricing is believed to be more efficient, which decreases the likelihood of finding evidence of predictability.[9]

When the samples are sorted according to different attributes of the stocks, we find an interesting pattern that the discrete events forecasting models are more likely to perform better among less risky stocks, e.g., using 5% as the significance level, our new models will beat the naive forecasting model significantly among 50% of largest stocks, 44% of stocks with lowest beta and 48% of stocks with lowest standard deviation. Such a finding indicates that the predictability of stock return events may not always be associated with more risky stocks or more illiquid stocks.

In a non-rigorous exercise,[10] for which we do not report the statistics, we find the following tendency in the functional form of the conditional probability function of (5.5). When we choose ARMA memory index coefficients in the range 0.2 to 0.4, the function form of (5.5) has relatively large function values in both end regions of the function domain and there is a slightly negative slope in the middle region.

5.2.2 Cross Sectional Method

In this part, the results are classified according to the round of out-of-sample forecasting and the size of learning samples.

First Round Out-of-Sample Forecasting with Small Sample Estimation

We present the forecasting evaluation results of first round out-of-sample forecasting in tabl5.2.

In table 5.2 panel A, we provide the statistics of the relative forecasting performance between the 30 competing models and the period-specific random walk model, i.e., $\Delta QPS_{i,m} = QPS_{i,0} - QPS_{i,m}$. Positive value of $\Delta QPS_{i,m}$ implies a better performance for our new models.

[9] The illiquidity in stock trading is a common arguement to explain the existence of predictability in finance literature.

[10] For each sample, we group the ARMA memory index of all the admissible stocks into 10 groups according to their value with an ascending order. We compute the average functional value of each group and investigate noticeable patterns on these group average values.

The data sample is drawn through the time varying cross sectional method. The probability function is estimated using one learning sample. We carry out both whole sample analysis (Panel A1) and sub sample analysis (Panel A2). The sub samples are sorted based on the market conditions prior to the forecasting periods.

For Panel A1,

(a) In the column 1-3, we present the p-value of the null hypothesis as in (5.9). Following the standard procedure in Hansen (2001), we provide the following information for each test result: the lower bound of the p-value, the consistent estimator of the p-value and the upper bound of the p-value.

(b) In the column 4-10, we present the average Quadratic Probability Score margin of the competing models. We group these statistics by both the resolution levels of wavelet models, i.e., low,medium and high and ARMA memory index coefficients, i.e., 0.1, 0.4, 0.7, 0.95.

(c) In the column 11-12, we present the best Quadratic Probability Score margin $\Delta QPS_{i,m}$ among all of the 30 competing models and the number of models that beat the reference model. As discussed in the section 2.6, we have four options for the regions of the function domain when we carry out out-of-sample forecasting. We present all of the results for different regions in the table.

For Panel A2, we provide same statistics for sub sample analysis for the comparison between the discrete event forecasting models and the reference model.

In table 5.2 Panel B, we presents the similar statistics to the Panel A for the comparison between the 30 discrete events forecasting models and the time invariant random walk model. The data sample is drawn through the time invariant cross sectional method. Again, whole sample analysis and sub sample analysis are carried out with sub samples classified according to the market condition prior to the forecasting periods.

In table Panel C, we presents the statistics for whole sample analysis when we compare the competing models with two reference models, a stock-specific random walk model and a period-specific random walk model. The data sample is drawn through the time invariant cross sectional method.

From Panels A and B, in both whole sample and sub sample analysis, there is little evidence that discrete events forecasting models will outperform the reference model. In most cases, Quadratic Probability Score (QPS) measures have negative value and the number of models that beat the reference model, if nonzero, is very small. The maximal number is 6 out of 30 models. The statistics of Superior Predictive Ability test further confirm the superiority of the reference model in each case.

From these two panels, we find that the performance of alternative models which are estimated using the time-varying sampling method is worse than those using time-invariant sampling method. In both sub sample and whole sample cases, on average, the models using the time-invariant sampling method has better relative performance in terms of the Quadratic Probability Score (QPS) and the number of models that beat the reference model. More convincing evidence comes from the Panel C. The models estimated using time-invariant sampling method outperform the reference model which corresponds to the time-varying case significantly. The p-value of the null hypothesis is 0, 0, 0 and all of the discrete events forecasting models have positive relative performance against that reference model.[11]

Based on the above observations, we draw the following conclusion. First, when the learning sample is relatively small, 1000 or so for each estimation, the forecasting performance of new alternative models are unsatisfactory. Second, the models estimated using the time-invariant sampling method is better than those using the time-varying sampling method in their competition with the reference model. One possible reason for inferior forecasting performance is a large estimation error due to the sampling error. As a way to reduce such errors, we expand the learning samples to check potential improvement. The results will be presented in the next section.

As we carry out the informal exercise to check the average functional form of the conditional probability function, we found some interesting regularities. For time varying cases, on average, when ARMA memory indexes have small coefficients, the conditional probability function is 'U' shaped. This shape implies that in a given month, if a stock has a relatively high or a relatively low value in ARMA memory indexes, the probability that it outperforms the market index in the next quarter increases. For time invariant cases, we find that when the ARMA memory index coefficients are small, there is a strong negative dependence between the value of ARMA

[11] Such outperformance does not happen to the models estimated using time-varying cross sectional methods when we compete them with the time invariant random walk models.

memory index and the conditional probability of a stock's outperforming index return. When the coefficients grow much bigger, i.e., 0.9 or 0.95, the dependence becomes positive.

First Round Out-of-Sample Forecasting with Large Sample Estimation

In this section, we will expand the learning samples for the estimation of the discrete events forecasting models. For each out-of-sample forecasting, we use 20 samples which are drawn through the time-invariant cross-sectional sampling method to estimate the conditional probability function (5.5).

In table 5.3 panel A and B, we provide the same statistics as in the table 5.2 panel A and panel B to show the relative forecasting performance between the 30 competing models and the time invariant random walk model.

The improvement of the performance is drastic. The evidence of predictability now becomes overwhelming through both sub sample and whole sample analysis. In the whole sample analysis, the p-value is uniformly 0 in all of the regions of the functional domain. For each region, the number of models that beat the reference model is 15,28,30,30 respectively. In addition, there is an interesting pattern for the optimal model in each forecasting region. For the forecasting regions that are close to the two ends, e.g., 10% end , 20% end, the models with larger ARMA memory index coefficients, i.e., 0.9 or 0.95, tend to outperform the reference model most. As the forecasting region expands to the whole domain, e.g., total, however, the models with medium ARMA memory index coefficients, i.e., 0.4, do the best forecasting job in comparison with the reference model. For the total region case, the relative performances of models are not monotone with the index coefficients. The models with the medium or extremely large indexes coefficients tend to be the better ones.

In the sub sample analysis, given that the market condition is strong bullish or strong bearish, the predictability of stock return events tends to be stronger. Both p-value statistics and the Quadratic Probability Score (QPS) measures lead to the same conclusion on such a pattern. Between the bull and bear market, the latter provides the stronger evidence of predictability.

As we draw the plot of the conditional probability functions in unreported exercises, we find that the models with a medium value in index coefficients, α, are downward sloped in most of the regions of the functional domain, which implies a price reversal phenomenon. As the ARMA memory index goes too large or too small, however, there is a reversal in the shape of the conditional probability function. The models with an extremely large value in index coefficients, i.e., 0.9 and 0.95, however, have an upward slope. This implies a momentum phenomenon. In previous studies, a momentum effect is hardly detectable using an explicit modeling approach. Within the framework of linear models, only negative serial correlation is unveiled. (see Lewellen 2002).

All of the above observations strongly suggest that nonlinearity happens to both the lag structure and the functional form of the conditional probability function (5.5). For the lag structure, when we condition on the combined past return performance of a stock, the momentum effect becomes unambiguous. This implies the significant impact from interaction items. For the functional form, the dependence between the conditional probabilities and the ARMA memory index is not always monotone. When the ARMA memory index comes to the end region, some reversal in the functional shape tends to happen. Such observation could be used to explain why models with small index coefficients are outperformed by the models with large index coefficients when we only choose the end regions to do forecasting.

Second Round Out-of-Sample Forecasting with Large Sample Estimation

In this section, we provide the statistics of the second round out-of-sample forecasting performance. The results are presented in the table 5.3 panel C. We do both whole sample analysis and sub sample analysis. The content of statistics is almost the same as in the panel B. The model we used to do forecasting for each sample is chosen based on the outcome of the first round out-of-sample forecasting (Table 5.3 Panel B).

For the whole sample analysis, all of the models yield positive relative performance. All but one regions show a significantly better performance for the discrete events forecasting models as compared with the time invariant random walk model.

For the sub sample analysis, all of the models still yield positive relative performance, which implies that, on average, the alternative models still perform better than the random walk model . However, except in the bearish market cases, the statistic significance becomes weak.

These results demonstrate that using an appropriate model evaluation approach, we are able to find a model that outperforms the random walk model significantly in both in-sample and out-of-sample cases, although caution needs to be taken due to the time variation in the strength of predictability. Moreover, the fact that the predictability is stronger under the bearish market condition is robust across different samples.

5.2.3 Summary

On the whole, based on the two-round out-of-sample forecasting evaluation scheme, we obtain persuasive evidence of the predictability of discrete stock events. Finding the source of such predictability is beyond the scope of this paper. However, based on the plot information, we find that the impact from price reversal and price momentum both play an evident role in determining such a predictability. The model estimation results and forecasting results give a clear idea about the complexity of the dependence between historical stock price information and the conditional probability functions. They also provide unambiguous evidences on the potential of the new forecasting modeling approach in handling complex modeling problems in probabilistic forecasting.

5.3 Conclusion

In this section, we propose a new econometric modelling approach to conduct probabilistic forecasting of discrete stock return events. This approach combines the ARMA memory index model with a semi-nonparametric estimation method, which adopts a wavelet basis as the basic building blocks for the functional form. Using the ARMA memory index approach, we do not need to truncate on conditioning variables in the modelling procedure. Using a wavelet-based semi-nonparametric estimation method, we do not have to set overly restrictive constraints on the form of the conditional expectation functions. Such combined flexibility allows us to capture the complex dependence between the conditional probability of the specific stock return event and the entire history of stock prices, which might be ignored in a standard parametric modelling approach due to model misspecification.

We test the potential of this new approach by applying it to study the predictability of stock return events such as an individual stock outperforming a benchmark stock index in monthly returns. We obtain the following interesting results .

(1) Using pure time series data, we find the evidence of predictability among individual stocks that have a long history.

(2) Using cross sectional data, we find that our conditional forecasting models can outperform the corresponding random walk model significantly when we use a two-round out-of-sample forecasting evaluation scheme.

(3) Both price reversal and price momentum phenomena are shown to exist explicitly using our modeling approach. The nonlinearity in both the lag structure and the conditional functional form is evident.

(4) The predictability of stock return events varies over time, in particular when the market conditions change.

Findings in this paper make contributions to the literature on stock return predictability and market efficiency in two aspects.

(1) With respect to the research methodology, our paper show that the discrete event forecasting modelling approach offers a convenient research tool to measure and investigate the predictability of stock returns. Not only does this method provide adequate modeling flexibility to detect various interesting predictable patterns in stock return series, but it also help reduce the concern about data mining due to the convenience of carrying out out-of-sample verification. Since both price reversal and price momentum phenomena are discovered convincingly using this modeling approach, it can be applied to analyze the cause of such predictability in a straightforward way.

(2) With respect to the empirical findings, this paper shows that the predictability of stock return events does exist. Past findings of price reversal and price momentum in individual stocks return are by no means spurious. In addition, the predictability can not be fully explained by the static risk factors since the discrete events forecasting models can easily outperform a stock specific random walk model. Finally, we find evidence of the time variation in such predictability, which is linked with the change of market conditions. Such a finding could also pave the way to a better understanding of the cause of the predictability in stock returns.

Table 5.1 Out-of-Sample Forecasting Performance for the Time Series Sampling Method

In table 5.1, we provide statistics of the out-of-sample forecasting performance of the Discrete-Event-Forecasting models in comparison with the reference model[1]. The data sample is drawn using the time series method.

The statistics include (a) the frequency of rejection of null hypothesis that the naïve model is the best model among the competing models. The p-value in use is consistent p_value estimator (See Hansen 2001). (b) mean (standard deviation) of the best QPS margin among all the samples that reject the null hypothesis. We also do sub sample analysis by classifying samples according to the rank in size, beta and standard deviation.

(Note: 1. Random Walk 1 model: use a stock specific probability.)

Sample		Fraction of samples in which the null hypothesis is rejected			Best QPS Margin
		Significance Level			Mean (Standard Deviation)
		5%	10%	15%	
Total Sample		0.405	0.46	0.535	0.00378 (0.002077)
Sub Samples Rank By Size	1 (Lowest)	0.44	0.5	0.58	0.00436 (0.002111)
	2	0.36	0.42	0.56	0.004405 (0.002051)
	3	0.32	0.4	0.44	0.002916 (0.001795)
	4 (Highest)	**0.5**	0.52	0.56	0.003408 (0.002062)
Sub Samples Rank By Beta	1 (Lowest)	0.4	0.42	0.54	0.003969 (0.001938)
	2	0.4	0.5	0.56	0.003458 (0.001933)
	3	0.38	0.44	0.52	0.004754 (0.002134)
	4 (Highest)	**0.44**	0.48	0.52	0.003093 (0.002061)
Sub Samples Rank By Standard Deviation	1 (Lowest)	0.4	0.44	0.62	0.003744 (0.00203)
	2	0.3	0.4	0.42	0.003957 (0.002553)
	3	0.44	0.52	0.58	0.004293 (0.002338)
	4 (Highest)	**0.48**	0.48	0.52	0.003047 (0.001742)

Table 5.2 Out-of-Sample Forecasting Performance for the Cross Sectional Sampling Method with Small Sample Estimation

In table 5.2, we provide statistics of the one-round out-of-sample forecasting performance of discrete events forecasting models in comparison with that of the reference models[1]. For each out-of-sample forecasting, we use one learning sample to estimate the conditional probability function. The data sample is drawn through the cross sectional method. We also choose different regions of the functional domain to do forecasting, e.g., a% end--the top and bottom a% value of ARMA memory index in each forecasting sample.

There are three groups of statistics. (a) The p-value of the null hypothesis that the reference model outperforms all the competing models with respect the expected forecasting error measure. There are three estimator for each p-value (1) Low: lower bound (2) Con: consistent estimator (3) Up: upper bound (see Hansen 2001) (b) the average QPS margin of competing models for different groups. (c) the best QPS margin and the best model among all the competing models and the total number of models that beat the reference model s.

Panel A presents the statistics for whole sample and sub sample analysis when we compare the competing models with the random walk 2 model. The data sample is drawn through the time varying cross sectional method. The sub samples are sorted based on the market conditions prior to the forecasting periods in out-of-sample forecasting.

Panel B presents the statistics for whole sample and sub sample analysis when we compare the competing models with the random walk 3 model. The data sample is drawn through the time invariant cross sectional method. The sub samples are sorted in the same way as above.

Panel C presents the statistics for whole sample analysis when we compare the competing models with the random walk 1 and random walk 2 model. Again, the data sample is drawn through the time invariant cross sectional method.

(Note: 1. Random Walk 1 model: use a stock specific probability; Random Walk 2 model: use a period-specific probability; Random Walk 3 model: use a time invariant probability)

Panel A1 Out-of-Sample Forecasting Results for the Time-Varying Cross Sectional Sampling Method—Total Sample Analysis

Forecasting Horizon	Sub-Set	P_Value for SPA			Average QPS Margin Grouped by Wavelets			Average QPS Margin Grouped by α				Best QPS Margin (Model)	Number of Better Models
		Low	Con	Up	High	Mid	Low	0.1	0.4	0.7	0.95		
Time-Varying	10% end	1	1	1	-0.00547	-0.00507	-0.00409	-0.00103	-0.0036	-0.00645	-0.00702	-0.00021	0/30
	20% end	1	1	1	-0.00405	-0.00371	-0.0031	-0.00131	-0.00233	-0.00431	-0.00567	-0.00108	0/30
	30% end	1	1	1	-0.00352	-0.00326	-0.00276	-0.00124	-0.00213	-0.0037	-0.00516	-0.00098	0/30
	Total	1	1	1	-0.0036	-0.00328	-0.00279	-0.0016	-0.0024	-0.0041	-0.00388	-0.0014	0/30

Panel A2 Out-of-Sample Forecasting Results for the Time-Varying Cross Sectional Sampling Method—Sub Sample Analysis

Forecast Horizon	Sub-Set	P-value											
		Sub-Sample 1 (Bear)			Sub-Sample 2			Sub-Sample 3			Sub-Sample 4(Bull)		
		Low	Con	Up	Low	Con	Up	Low	Con	Up	Low	Con	Up
One-month	10% end	0.588	0.657	0.863	1	1	1	0.163	0.183	0.289	0.521	0.521	0.954
	20% end	1	1	1	1	1	1	0.248	0.263	0.439	1	1	1
	30% end	1	1	1	1	1	1	0.271	0.307	0.488	1	1	1
	Total	1	1	1	1	1	1	1	1	1	0.347	0.42	0.665

Forecast Horizon	Sub-Set	Other Performance Statistics							
		Sub-Sample 1 (Bear)		Sub-Sample 2		Sub-Sample 3		Sub-Sample 4(Bull)	
		Best QPS Margin	Number of Better Models	Best QPS Margin	Number of Better Models	Best QPS Margin	Number of Better Models	Best QPS Margin	Number of Better Models
One-month	10% end	0.000137	1/30	-0.00024	0/30	0.004492	14/30	0.000182	1/30
	20% end	-0.00083	0/30	-0.00035	0/30	0.002726	16/30	-0.00083	0/30
	30% end	-0.00078	0/30	-0.00113	0/30	0.001892	13/30	-0.00047	0/30
	Total	-0.00149	0/30	-0.00217	0/30	-0.00045	0/30	0.001751	2/30

Panel B1 Out-of-Sample Forecasting Results for the Time-Invariant Cross Sectional Sampling Method—Total Sample Analysis

Forecasting Horizon	Sub-Set	P_Value			Average QPS Margin Grouped by Wavelets			Average QPS Margin Grouped by α				Best QPS Margin (Model)	Number of Better Models
		Low	Con	Up	High	Mid	Low	0.1	0.4	0.7	0.95		
One-month	10% end	0.226	0.327	0.574	-0.00282	-0.00105	3.55E-05	-0.0019	-0.00109	-0.00107	-0.00065	0.000592	5/30
	20% end	0.339	0.439	0.73	-0.00169	-0.00069	-3.35E-05	-0.00154	-0.00061	-0.00063	-0.00037	0.000272	4/30
	30% end	0.644	0.802	0.962	-0.00136	-0.00059	-0.00012	-0.00133	-0.00064	-0.00046	-0.00034	2.08E-05	2/30
	Total	1	1	1	-0.00148	-0.00071	-0.00025	-0.00123	-0.00078	-0.00073	-0.00032	-0.00013	0/30

Panel B2 Out-of-Sample Forecasting Results for the Time-Invariant Cross Sectional Sampling Method—Sub Sample Analysis

Forecast Horizon	Sub-Set	P-value											
		Sub-Sample 1 (Bear)			Sub-Sample 2			Sub-Sample 3			Sub-Sample 4(Bull)		
		Low	Con	Up	Low	Con	Up	Low	Con	Up	Low	Con	Up
One-month	10% end	0.503	0.597	0.82	0.252	0.286	0.541	0.738	0.76	0.922	0.373	0.519	0.743
	20% end	0.55	0.673	0.885	0.21	0.267	0.509	0.409	0.435	0.727	0.507	0.637	0.764
	30% end	0.551	0.619	0.929	0.485	0.526	0.878	0.454	0.551	0.875	0.613	0.636	0.819
	Total	1	1	1	0.499	0.54	0.936	0.346	0.346	0.811	0.556	0.616	0.782

Forecast Horizon	Sub-Set	Other Performance Statistics							
		Sub-Sample 1 (Bear)		Sub-Sample 2		Sub-Sample 3		Sub-Sample 4(Bull)	
		Best QPS Margin	Number of Better Models	Best QPS Margin	Number of Better Models	Best QPS Margin	Number of Better Models	Best QPS Margin	Number of Better Models
One-month	10% end	0.000834	6/30	0.001092	6/30	0.000192	3/30	0.000646	5/30
	20% end	0.000287	5/30	0.000718	8/30	0.000381	3/30	0.000323	4/30
	30% end	0.000136	4/30	0.000167	5/30	0.000254	3/30	0.000231	6/30
	Total	-0.00016	0/30	6.04E-05	3/30	0.000164	3/30	0.000201	5/30

Panel C Out-of-Sample Forecasting Results for the Time-Invariant Cross Sectional Sampling Method against Random Walk 1 ,2

	Reference Model	P_value			Average QPS Margin Grouped by Wavelets			Average QPS Margin Grouped by α				Best QPS Margin	Number of Better Models
		Low	Con	Up	High	Mid	Low	0.1	0.4	0.7	0.95		
One-Month Ahead	Random Walk 1	0.006	0.006	0.006	-0.00026	0.000518	0.000976	-1.01E-05	0.000442	0.000491	0.000904	0.001092	23/30
	Random Walk 2	0	0	0	0.00672	0.007494	0.007952	0.006966	0.007418	0.007467	0.00788	0.008068	30/30

Table 5.3 Out-of-Sample Forecasting Performance for the Time Invariant Cross Sectional Sampling Method with Large Sample Estimation

In table 5.3, we provide statistics of the two-round out-of-sample forecasting performances of discrete events forecasting models in comparison with the Random Walk 3 model[1]. The data sample is drawn through time invariant cross sectional method. For each out-of-sample forecasting, we use 20 learning samples to estimate the conditional probability function. Also, we chose different types of regions of the functional domain to do out-of-sample forecasting, e.g., a% end refers to the top and bottom a% value of ARMA memory index in each forecasting sample.

There are three groups of statistics. (a) The p-value of the null hypothesis that the reference model outperforms all the competing models with respect to the expected forecasting error measure. There are three estimator for each p-value (1) Low: lower bound (2) Con: consistent estimator (3) Up: upper bound (see Hansen 2001) (b) the average QPS margin of competing models for each specific groups. (c) the best QPS margin and the best model among all the competing models and the total number of models that beat the reference model s.

Panel A presents the statistics for whole sample analysis when we compare the competing models with the reference model in the first-round out-of-sample forecasting.

Panel B presents the statistics for sub sample analysis in first round out-of-sample forecasting. The sub samples are sorted based on the market conditions prior to the forecasting periods.

Panel C presents the statistics for whole sample and sub sample analysis when we compare the competing models with the random walk 3 model in the second round out-of-sample forecasting.

(Note: The reference model uses a time invariant constant probability to make forecasting.)

Panel A The First Round Out-of-Sample Forecasting Performance Against the Random Walk 3 model —Total Sample Analysis

Forecast Horizon	Sub-Set	P_value			Average QPS Margin Grouped by Wavelets			Average QPS Margin Grouped by α				Best QPS margin (Model)	Number of better Model
		Low	Con	Up	High	Mid	Low	0.1	0.4	0.7	0.95		
One-month	10% end	0	0	0	-2.65E-05	0.000131	0.000188	-0.00091	-0.00031	0.000501	0.001109	0.001305 (27)	15/30
	20% end	0	0	0	0.000333	0.000484	0.000537	-1.40E-05	0.000265	0.000604	0.00086	0.001007 (27)	28/30
	30% end	0	0	0	0.000503	0.000642	0.000686	0.00038	0.000565	0.000693	0.000779	0.000847 (30)	30/30
	Total	0	0	0	0.000623	0.000769	0.000824	0.000648	**0.00082**	0.000686	**0.000763**	0.000916 (12)	30/30

Panel B The First Round Out-of-Sample Forecasting Performance against the Random Walk 3 model—Sub Sample Analysis

Forecast Horizon	Sub-Set	P-value of Null Hypothesis											
		Sub-Sample 1 (Bear)			Sub-Sample 2			Sub-Sample 3			Sub-Sample 4(Bull)		
		Low	Con	Up	Low	Con	Up	Low	Con	Up	Low	Con	Up
One-month	10% end	**0**	**0**	**0**	0.073	0.086	0.125	0.043	0.044	0.057	**0**	**0**	**0**
	20% end	**0**	**0**	**0**	0.056	0.056	0.072	0.035	0.041	0.043	**0.003**	**0.003**	**0.005**
	30% end	**0**	**0**	**0**	0.041	0.046	0.046	0.138	0.158	0.159	**0.013**	**0.013**	**0.013**
	Total	**0.001**	**0.001**	**0.001**	0.096	0.096	0.096	0.187	0.22	0.223	**0.001**	**0.001**	**0.001**
		Sub-Sample 1 (Bear)			Sub-Sample 2			Sub-Sample 3			Sub-Sample 4(Bull)		
		Best QPS margin	Best Model	No of Better Model	Best QPS margin	Best Model	No of Better Model	Best QPS margin	Best Model	No of Better Model	Best QPS margin	Best Model	No of Better Model
One-Month	10% end	**0.002106**	**27**	**30/30**	0.000965	27	7/30	0.000949	27	15/30	**0.001921**	**30**	**11**/30
	20% end	**0.001806**	**24**	**30/30**	0.000607	30	23/30	0.000794	27	21/30	**0.001099**	**30**	**11**/30
	30% end	**0.001583**	**24**	**30/30**	0.000595	21	27/30	0.000533	30	22/30	**0.000949**	**30**	**30**/30
	Total	**0.001443**	**30**	**30/30**	0.000747	12	30/30	0.000399	30	14/30	**0.001533**	**9**	**30**/30

Panel C The Second-Round Out-of-Sample Forecasting Performance against the Random Walk 3 Model

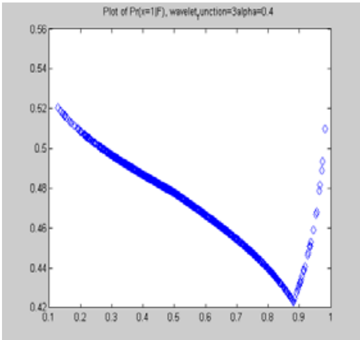
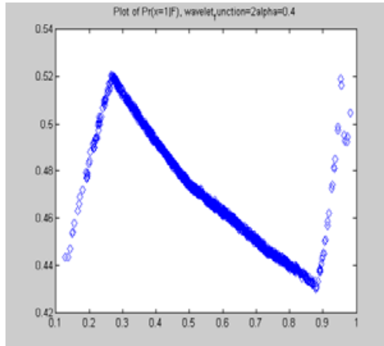
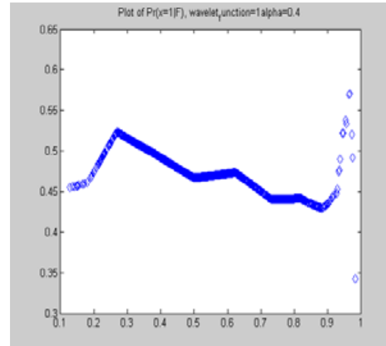
Forecast Horizon	Sub-Set	Sub Sample Performance Measure											
		Sub-Sample 1 (Bear)			Sub-Sample 2			Sub-Sample 3			Sub-Sample 4(Bull)		
		QPS margin	Std Error	Model In Use	QPS margin	Std Error	Model In Use	QPS margin	Std Error	Model In Use	QPS margin	Std Error	Model In Use
One-month	10% end	**0.001262**	**0.000721**	27	0.000905	0.000536	27	0.000383	0.000591	27	0.000288	0.000540	30
	20% end	**0.001119**	**0.000567**	24	0.000428	0.00032	30	0.000484	0.000387	27	0.000325	0.000424	30
	30% end	**0.000694**	**0.00047**	24	0.000575	0.000319	21	0.000432	0.000460	30	0.000372	0.000412	30
	Total	**0.0003**	**0.000447**	30	0.000466	0.000359	12	0.000355	0.000393	30	0.000588	0.000496	9

		Total Sample		
		QPS Margin	Std Error	Model In Use
One-month	10% end	**0.0007**	**0.000314**	**27**
	20% end	**0.000527**	**0.000215**	**27**
	30% end	**0.000372**	**0.000206**	**30**
	Total	**0.000551**	**0.000244**	**12**

Figure 5.1 Plots of Estimation and Forecasting Results of the Discrete-Event-Forecasting Model for Stock Return Forecasts

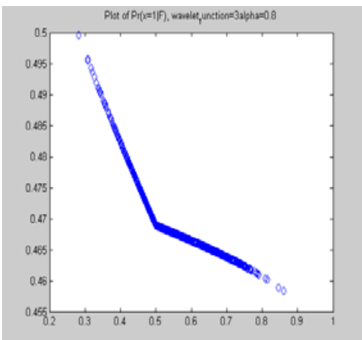
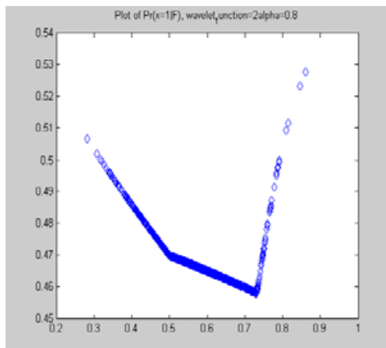
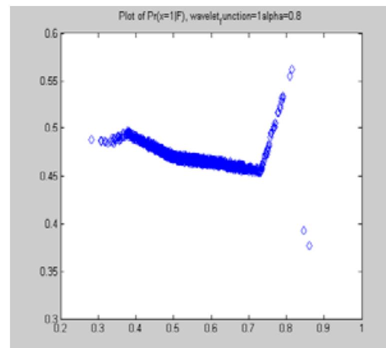
α=0.4

Low Resolution	Medium Resolution	High Resolution
12	11	10

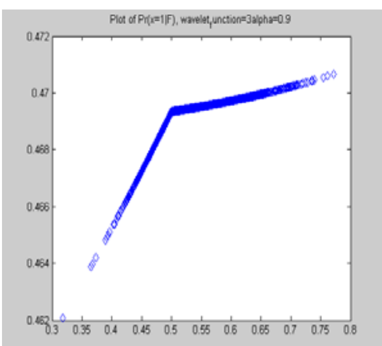
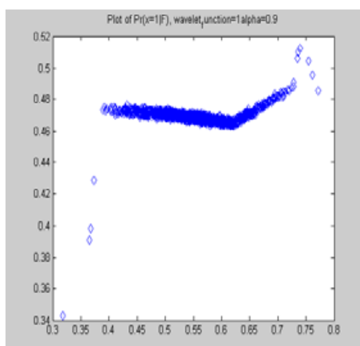
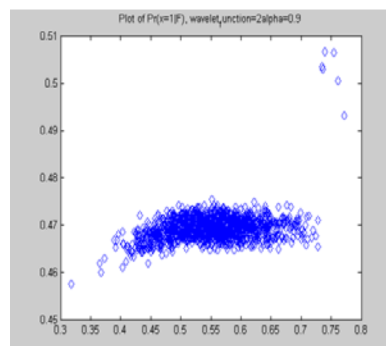
α=0.8

24	23	22

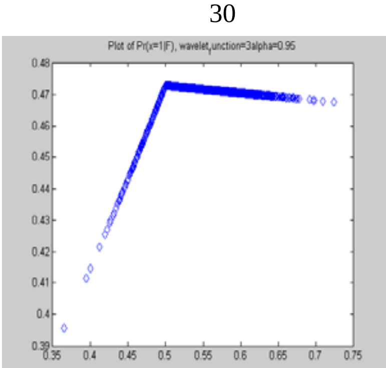
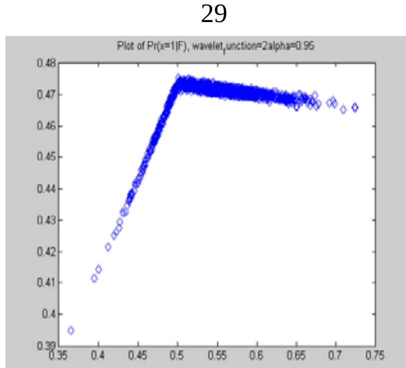
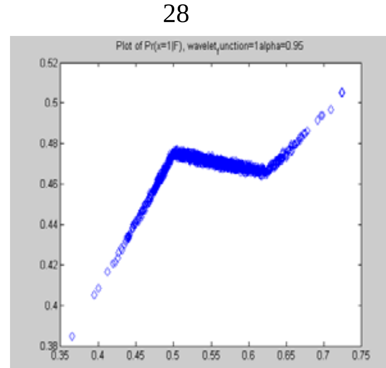
α=0.9

27	26	25

α=0.95

30	29	28

Basics of Functional Analysis

Basic Definitions:

In the nonparametric and semi-nonparametric estimation, people try to estimate an unknown functional form with minimal prior knowledge imposed. Therefore, the object under estimation is the whole structure of functional form rather than an unknown scalar or vector in the traditional parametric estimation. The first problem is how to quantitatively describe the relationship between the structure estimator and the true structure. In parametric estimation, we can simply adopt the distance formula for the Euclidean space to fulfill such a goal. In nonparametric and semi-nonparametric cases, we need to set up new rules in order to implement this task.

Normed Linear Space and Banach Space

Definition 2 ***Linear Vector Space***

A linear vector space W over $\mathbb{R}$ is nonempty set W with a mapping: $(x_1,x_2) \to x_1 \oplus x_2$, from $W \times W$ into W,which we call addition and a mapping: $(\alpha, x) \to \alpha x$,from $\mathcal{F} \times W$ into W which we call scalar multiplication.

Briefly, the linear vector space is a set of elements that is closed under addition and scalar multiplication. Note that Euclidean space $\mathbb{R}^n$, $n \in \mathbb{Z}$ is just a special case of linear vector space. It is desirable that space of functional form is closed under these two basic algebra operations. This makes it possible to extend a lot important results developed in real analysis to functional analysis.

Definition 3 ***Norm***

A norm is a non-negative set function on a linear vector space, $\|\cdot\| : W \to R^+$ such that

(1) $\|x\| = 0$, *iff* $x = 0$;

(2) $\|x \oplus y\| \leqslant \|x\| + \|y\|$, *for* $x, y \in W$.*(Triangle Property)*;

(3) $\|\alpha x\| = |\alpha| \|x\|$, *for all* $x \in W$ *and* $\alpha \in \mathcal{F}$.

It is apparent that the way to compute the distance between elements and origin in $\mathbb{R}^n$ space just satisfies the definition of norm. Therefore, norm can be considered as a generalized distance between elements in a linear vector space.

Definition 4 ***Normed Linear Vector Space***

A normed linear space is linear vector space X with a norm $\|\cdot\|$ on it and is denoted by $(X, \|\cdot\|)$.

Definition 5 ***Complete Normed Linear Vector Space***

A normed linear vector space is complete if any Cauchy sequence in the space converge to element of this space. A complete normal linear vector space is also called Banach space.

(Note: the Cauchy Sequence is defined as follows.

Let $\{a_j\}_{j=1}^{\infty}$be a sequence of real (or complex) numbers. We say that the sequence satisfies the Cauchy criterion (or simply is Cauchy) if for each $\varepsilon > 0$ there is an integer $N > 0$ such that if $j, k > N$ then

$$|a_j - a_k| < \varepsilon$$

)

In most semi-nonparametric estimation, we would like to use a complete space as the parameter space. The benefit of such an assumption will be seen in the later section.

Inner Product Space and Hilbert Space

In this part, we will introduce a special case of Banach space, which is extremely important to semi-nonparametric estimation via orthogonal series expansion.

Definition 6 ***Inner Product***

A inner product is a mapping $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$, satisfying that for $x, y \in$ vector space X

(Conjugate symmetry) $\langle x, y \rangle = \langle y, x \rangle^*$

(Positive definite) $\langle x, x \rangle \geqslant 0$ and $\langle x, x \rangle = 0$ iff $x = 0$.

(Linear in the first argument) $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$

Here, X is vector space and R is the set of complex number.

An example of inner product space is $L_2[a, b]$ of measurable function on $[a, b]$ with the inner product $\langle x, y \rangle = \int_a^b x(t)y(t)dt$ for $x, y \in L_2[a, b]$.

Definition 7 ***Inner Product Space***

A vector space endowed with inner product is called inner product space.

For an inner product space, the norm is usually defined as $\|x\| = \sqrt{\langle x, x\rangle}$. An example of inner product space is $L_2[\mathbf{a}, \mathbf{b}]$ of measurable functions on $[a, b]$ with the inner product $\langle x, y\rangle = \int_a^b x(t)y(t)^* dt$, for $x, y \in L_2[\mathbf{a}, \mathbf{b}]$. For each $x \in L_2[\mathbf{a}, \mathbf{b}]$, $\int_a^b [x(t)]^2 dt < \infty$.

Definition 8 ***Hilbert Space***

A complete inner product space is called Hilbert Space.

Orthonormal Basis In Inner Product Space

Now we'll introduce the core concepts concerning orthogonal series.

Definition 9 ***Orthogonality***

$x, y \in X$, an inner product space, is orthogonal to each other if $\langle x, y\rangle = 0$. We denote it as $x \perp y$.

Definition 10 ***Orthogonal set/sequence and Orthonormal set***

A subset S of an inner product space X is called an orthogonal set if $x \perp y$ for distinct $x, y \in X$. If S is denumerable, we say that S is an orthogonal sequence. Furthermore, if $\langle x, x\rangle = 1$ for all $x \in S$, we call it orthonormal set.

Definition 11 ***Orthonormal base***

An orthonormal set is called orthonormal base for inner product space X if the set isn't properly contained in any other orthonormal set. It is also called maximal orthonormal set or complete orthonormal set.

Properties of Hilbert Space

Theorem 12 *An orthonormal sequence (x_n) in a Hilbert space X is an orthonormal base if and only if any one of the equivalent conditions below is satisfied:*

(a) $cl[(x_n)] = X$;

(b) $[(x_n)]^\perp = \{0\}$.

(c) $\|x\|^2 = \sum_{n\in\mathbf{N}} |\langle x, x_n\rangle|^2$ holds for all $x \in X$.

(d) $x = \sum_{n\in\mathbf{N}} \langle x, x_n\rangle x_n$ for every $x \in X$.

From the above theorem, we know that in order that Hilbert space X is norm isomorphic to space l^2, X must be the closure of space spanned by its orthonormal base. Another way to find such a space is as follows:

Theorem 13 *An inner product space is separable if and only if it has a complete orthonormal sequence* (x_n)*.Furthermore, in a separable inner product space, any* $x \in X$ *can be written uniquely in the form* $x = \sum_{n \in \mathbf{N}} \langle x, x_n \rangle x_n$ *for any orthonormal basis* (x_n)*.*

Fourier Analysis

Convergence Properties of Fourier Series

Convergence Mode

Definition 14 ***Mean Convergence***

Let (x_n) *be a sequence in* $L_2[a,b]$*,* x_n *converges to* x *in the mean if* $\int_a^b (x_n(t) - x(t))^2 dt \to 0$ *as* $n \to \infty$*.*

Definition 15 ***Pointwise Convergence***

A sequence (x_n) *of real or complex valued function with a common domain* T *converges pointwise to the function* x *if* $x_n(t) \to x(t)$ *as* $n \to \infty$ *for all* $t \in T$*.*

Definition 16 ***Uniform Convergence***

Convergence with respect to sup norm $\|\cdot\|_\infty$ *is called uniform convergence, where* $\|\cdot\|_\infty$ *is expressed as* $\|x\|_\infty = \sup_{t \in T}\{|x(t)| : t \in T\}$

From the results in last section, Fourier series $s_N = \sum_{n \le N} \langle x, x_n \rangle x_n$ will converge to x in the mean for any $x \in L_2[a,b]$. It is natural to ask under what condition Fourier series could converge in a stronger sense, i.e. pointwise convergence or uniform convergence. In the following part, several famous theorems are given in order to answer these questions.

Convergence Properties of Trigonometric Fourier Series

In last several sections, we define the Fourier series in a very general sense, that is, $s_N = \sum_{n \le N} \langle x, x_n \rangle x_n$, where $\{x_n\}$ is orthonormal basis for inner product space with inner product $\langle \cdot, \cdot \rangle$. From now on, we'll focus our discussion on the Fourier series with specific form:

$$s_k^1(t) = \sum_{n=-k}^{k} c_n e^{-int}, \tag{5.10}$$

and

$$s_k^2(t) = a_0 + \sum_{n=1}^{k} a_n \cos nt + \sum_{n=1}^{k} b_n \sin nt. \tag{5.11}$$

The $s_k^1(t)$ is called trigonometric polynomials and $s_k^2(t)$ is called trigonometric series. In order to be Fourier series, $c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} x(t)e^{-int}dt$ and $a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} x(t)\cos(nt)dt$, $b_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} x(t)\sin(nt)dt$.It is easy to verify that these two series are convertible to each other. Therefore, if $s_k^1(t)$converges as $k \to \infty$, so does $s_k^2(t)$.

Before we introduce the following theorems, we need to introduce several concepts.

Definition 17 ***Periodic Extension of Function***

Function f is $(b-a)$-periodic extension of function g on $[a, b]$ if

$$\begin{aligned} f(t) &= g(t) \text{ if } t \in [a, b] \\ &= g(t') \text{ if } t = t' + k(b-a) \text{for some } \quad k \in Z \text{ and } t' \in [a, b] \end{aligned}$$

Definition 18 ***One-Sided Derivatives***

Let f be defined in an interval contain c. A right-handed derivative $f_r'(c)$ of f at a point c is

$$f_r^{'}(c) = \lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h}$$

If point c is a discontinuous point with $f(c^+) = \lim_{h \to 0^+} f(c+h)$ and $f(c^-) = \lim_{h \to 0^-} f(c+h)$, then right-handed derivative is expressed as

$$f^{'}(c^+) = \lim_{h \to 0^+} \frac{f(c+h) - f(c^+)}{h}$$

Similarly, we can get definition of $f'(c^-)$.

Definition 19 ***Piecewise Smooth Function***

Function f is called piecewise smooth function if f is composed of a finite differentiable components and $f'(c^+)$ and $f'(c^-)$ exist at every point $c \in [a, b]$,except the endpoints where only $f'(a^+)$ and $f'(b^-)$ exist.

Definition 20 ***Bounded Variation***

Let h be a complex-valued function defined on the (finite) closed interval $[a, b]$ and let $P = \{t_0, t_1, t_2, ...t_n\}, t_0 = a < t_1 < ... < t_n = b$, be a partition of $[a, b]$. The variation of h with respect to P is

$$V(h, a, b) = \sum_{i=0}^{n-1} |h(t_{i+1}) - h(t_i)|$$

The (Total) variation of h over $[a, b]$ is

$$V(h, a, b) = V(h) = \sup_p V_p$$

as P ranges over all partition of $[a, b]$.h is of bounded (or finite) variation (call BV later) on $[a, b]$ if $V(h) < \infty$.

The next theorem concerns the pointwise convergence of Fourier series. It is first completed by Dirichlet.

Theorem 21 ***Pointwise Convergence of Trigonometric series (Dirichlet)***

Let f be the $2\pi-$periodic extension of an integrable function on $[-\pi, \pi]$.If $f'(t^+)$ and $f'(t^-)$ exist everywhere-as happens for any piecewise smooth function f– then the Fourier series (equation 5.11) for f converges to $\frac{1}{2}(f(t^-) + f(t^+))$ at every t; at points t of continuity the series there converge to $f(t)$.

In this theorem, the function class under consideration is $2\pi-$periodic extension of a piecewise smooth function. It may seems quite restrictive condition. However, as long as the function f **under study is piecewise smooth function with a compact support $[a, b]$, simple algebra manipulation can make it a function f^*defined on $[-\pi, \pi]$.Then we can extend this function by $f^{**}(t + 2\pi) = f^*(t)$ to get it into the function class we need. Also, it is noted that the class of piecewise smooth function on a compact support is sufficiently large to fit the demand of most social science study. But still, restriction on piecewise smooth can be relaxed further.

Theorem 22 ***Pointwise Convergence of Trigonometric series (Jordan)***

If f is the $2\pi-$periodic extension of function of BV on $[-\pi,\pi]$ which implies that f is integrable on $[-\pi,\pi]$), then for any t, its Fourier Series converges (equation 5.11)to $\frac{1}{2}(f(t^-)+f(t^+))$.

The above two theorems not only give the sufficient condition for convergence of Fourier series but also point out where each point converges. They offer a real breakthrough in Fourier analysis. There are still a lot of other important theorems concerning the property of Fourier Series. Here we won't list them. Interested readers could refer to George Bachman, et al (2000) for a brief summary on all these results.

Dirichlet Kernel and Kernel Estimation

In the last part, we give quite a few important results concerning the convergence properties of Fourier Series. All these efforts aim at justifying the possibility of using $s_\infty(t) = \sum_{n=-\infty}^{\infty}\frac{1}{2\pi}\int_{-\pi}\pi f(t)e^{-inx}dxe^{-int}$ to represent the targeted function f . In this part, we would like to look a little deeper to see why Fourier Series, especially the Trigonometric Series, possess these convergence property (pointwise or uniformly). We will first introduce Dirichlet Kernel which play a decisive role in Fourier Series.

The nth Dirichlet Kernel is computed as follows:

$$\begin{aligned} D_n(t) = \sum_{k=-n}^{n} e^{ikt} &= 1 && n=0 \\ &= 1+2\textstyle\sum_{k=1}^{n}\cos kt && n\in\mathbf{N} \end{aligned} \tag{5.12}$$

By calculating the summation, we have much more compact expression form

$$\begin{aligned} D_n(t) &= \tfrac{\sin(n+1/2)t}{\sin(t/2)} && t\neq 0,\pm 2\pi,\pm 4\pi,... && n\in\mathbf{N}\cup\{0\} \\ &= 2n+1 && && \text{otherwise} \end{aligned} \tag{5.13}$$

It is easy to see

$$s_N(t) = \sum_{n=-N}^{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-inx}dxe^{-int} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_n(t+x)dx$$

The following theorems concerns an extremely important property of Dirichlet Kernel which has direct implication of the property of Fourier series S_N.

Theorem 23 ***Riemann-Lebesgue Lemma*** *For any* $f \in L_1^r[a,b]$, $-\infty \leq a \leq b \leq \infty$, *and any bounded measurable function* h *defined on* $\mathbb{R}$, *it follows that*

$$\lim_{c \to \pm\infty} \frac{1}{c} \int_0^c h(t)dt \to 0$$

then

$$\lim_{w \to \infty} \int_a^b f(t)h(wt)dt = 0$$

Theorem 24 ***Riemann-Lebesgue Property of*** $\mathbf{D}_n$

For any $f \in L_1^r[-\pi,\pi]$, *and* $r \in (0,\pi]$,

$$\lim_n \int_r^\pi f(t)D_n(t)dt = 0 \tag{5.14}$$

where D_n *is Dirichlet Kernel. (see 5.13)*

Then combining the above result with the easy fact that

$$\int_0^\pi D_n(t)dt = \pi \text{ for all } n \in \mathbb{N} \cup \{\mathbf{0}\}.$$

we will find that $D_n(t)$ has such an interesting property: the total area under the curve $D_n(t)$ for $[0,\pi]$ is fixed as π. But if we fix $r \in (0,\pi]$, given a sufficient large n, we have the total area under the curve for $[r,\pi]$ getting arbitrarily close to zero (see 5.14. Since r is arbitrary, it is easy to visualize that as $n \to \infty, D_n(t)$ serves like Dyadic-delta function δ_0 which is defined as follows.

Definition 25 ***Dyadic-Delta Function:***

Dirac Delta Function is a function δ_x satisfying:

$$\delta_x(t) = 0 if t \neq x$$

and

$$\int \delta_x(t)dt = 1$$

If so, as long as $f \in L_1^r[-\pi, \pi]$ is well defined at point 0 and it is continuous from right, it is always true that .

$$\lim_{n\to\infty} \int_0^{\pi} f(t)D_n(t)dt = \pi f(0^+) \tag{5.15}$$

This result directly relate with convergence property of Fourier series[12] .

From the above deposition, we now have a clear idea of how the size of series expansion affects the goodness of approximation:

Under mild regularity conditions, for each point $x \in supp(f)$, one can always find a small neighborhood U_x such that the value of function won't change too much within U_x. As a result, the weighted average over U_x can well approximate $f(x)$.On the other hand, due to localization property, if the expansion size N is large enough, $\int_0 \pi f(t)D_n(t)dt$ just yield the weighted average over U_x.Therefore, the size of expansion will determine how well the approximation can be.

[12] As long as we can formulate the Fouries Series in the following way

$$s_n = \int_0^{\pi} f(t+u)D_n(t)dt$$

then the pointwise convergence can be achieive in the spirit of equation 5.15. The derivation is as follows:

$$\begin{aligned}
s_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx + \frac{1}{\pi}\left[\sum_{k=1}^{n}\cos(kt)\int_{-\pi}^{\pi} f(x)\cos(kx)dx + \sin(kt)\int_{-\pi}^{\pi} f(x)\sin(kx)dx\right] \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\left[\frac{1}{2} + \sum_{k=1}^{n}(\cos(kt)\cos(kx) + \sin(kt)\sin(kx))\right]dx \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\left[\frac{1}{2} + \sum_{k=1}^{n}\cos k(t-x)\right]dx \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)D_n(x-t)dx = \frac{1}{\pi}\int_{-\pi-t}^{\pi-t} f(t+x)D_n(x)dx \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t+x)D_n(x)dx = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t-x)D_n(x)dx \qquad (*) \\
&= \frac{1}{\pi}\int_{-\pi}^{0} f(t+x)D_n(x)dx + \int_0^{\pi} f(t+x)D_n(x)dx \\
&= \frac{1}{2\pi}\int_0^{\pi}(f(t-x) + f(t+x))D_n(x)dx
\end{aligned}$$

From this fact, we can deduce the following regularity as to the convergence of Fourier series: the smoother the original function is,the fewer number of coefficients is required for a good approximation by Fourier series and the more quickly the sequence Fourier coefficients will dampen.

The method of using Kernel function like Dirichlet Kernel to approximate target function $f(\cdot)$ is by no means exclusive attributes of Fourier analysis. It is a widely popular technique in engineering and mathematics analysis. Since the kernel $K(\cdot)$can be used to reproduce a function in the sense

$$f(x) = \int f(t)K(t,x)dt$$

it is also called reproducing kernel.

Wavelet Analysis

Multi-Resoluation Anlysis (MRA)

Basic Definition

Multiresolution analysis consists of a sequence of subspace V_j , $j \in Z$, of $\mathbb{L}^2(\mathbb{R})$ satisfying:

1. $V_j \subset V_{j+1}$, for $j \in Z$
2. $f \in V_j$, if and only if $f(2(\cdot)) \in V_{j+1}$
3. $\bigcap\limits_{j\in Z} V_j = \{0\}$;
4. $cl(\bigcup\limits_{j\in Z} V_j) = L^2(R)$;
5. There exist a function $\varphi \in V_0$, such that $\{\varphi(\cdot - k) : k \in Z\}$ is orthonormal basis for $V_0.\varphi(\cdot)$ is called scaling function or father wavelet function.

With these five conditions satisfied, we can explicitly construct an orthonormal wavelet basis $\{\psi_{j,k;}j,k \in Z\}$ of $\mathbb{L}^2(\mathbb{R}), \psi_{j,k;} = 2^{-\frac{j}{2}}\psi(2^{-j}x - k)$,such that for all f in $\mathbb{L}^2(\mathbb{R})$,

$$P_{j+1}f = P_jf + \sum_{k\in Z} \langle f, \psi_{j,k}\rangle \psi_{j,k} \tag{5.16}$$

where P_j is the orthogonal projection onto V_j.

Before we show how to get wavelet basis under these five conditions, we want to point out all the above conditions are not independent with each other and some relaxation can be made.

Theorem 26 *Conditions (1) , (2) and (5) in definition of MRA imply (3). This is the case even if in (5), we only assume that , $\{\varphi(\cdot - k) : k \in Z\}$ is a Riesz basis*[13].

Theorem 28 *Let $\{V_j : j \in Z\}$ be a sequence of closed subspaces of $L^2(R)$ satisfying (1) (2) (5) ; assume that the scaling function φ of condition (5) is such that $|\hat{\varphi}|$[14] is continuous at 0. Then, the following two conditions are equivalent:*

(i) $\hat{\varphi}(0) \neq 0$ *(ii)* $cl(\bigcup_{j \in Z} V_j) = L^2(R)$;

Theorem 29 *For a Riesz basis for V_0, defined as above, we can then find $\gamma \in V_0$, such that $\{\varphi(\cdot - k) : k \in Z\}$ is an orthonormal basis for V_0.*

See Daubechiez[1992] Chapter 5 and Hernandez &Weiss[1996], Chapter 2 for the proof of above theorems. From above three theorems, we'll recognize that the conditions (1), (2) and relaxed (5) in the definition of MRA are the fundamental ones. It is interesting to find that the first two properties is consistent with our intuitive explanation of MRA in last section. The increasing sequence V_j is just the sequence of fluctuation components. By condition (2), we find the fluctuation component possess higher resolution as j increase. This is because the functions in V_{t+1}can complete the same degree of variation in half-sized region as their counterpart in V_t. As a result, function in this space can take on a more volatile pattern.

In the next section, we will give the expression of wavelet function in terms of $\varphi(\cdot)$.

---

[13] ***Riesz basis***

For every $f \in V_0$, there exists a unique sequence$\{\alpha_n\}_{n \in Z} \in l^2(Z)$ such that

$$f(x) = \sum_{n \in Z} \alpha_n \varphi(x - n)$$

with convergence in $L^2(R)$ and

$$A \sum_{n \in Z} |\alpha_n|^2 \leq \left\| \sum_{n \in Z} \alpha_n \varphi(x - n) \right\|_2^2 \leq B \sum_{n \in Z} |\alpha_n|^2$$

Definition 27 *with $0 < A \leq B < \infty$ constants independent of $f \in V_0$.Then $\varphi(\cdot - n)$ is Riesz Basis of space $L^2(R)$. We can $\varphi(\cdot)$ as a scaling function.*

[14] $\hat{\varphi}(w) = \frac{\pi}{2} \int_{-\pi}^{\pi} \varphi(t) e^{-itw} dt$ *is fourier transform of original function $\varphi(t)$.*

General Procedure of Wavelets Construction

Since $V_j \subset V_{j+1}$, we can define a subspace W_j, such that $W_j \perp V_j$, and $V_{j+1} = \{x : x = y + z, y \in W_j, z \in V_j\}$, briefly $V_{j+1} = W_j \oplus V_j.W_j$ can be called orthogonal complement of V_j in V_{j+1}. It is easy to see $W_{j'} \subset V_j$, if $j > j'$. So $W_{j'} \perp W_j$, if $j \neq j'$.

Then, we can write $V_j = W_{j-1} \oplus V_{j-1}$into $V_j = W_{j-1} \oplus W_{j-2} \oplus V_{j-2} = \bigoplus_{k=-\infty}^{j-1} W_k$. Then we know

$$L^2(R) = \bigoplus_{k=-\infty}^{\infty} W_k, \tag{5.17}$$

since $V_j \subset V_{j+1}$ and $\lim_{J\to\infty} \bigcup_{j\le J} V_j = L^2(R)$..

The result shown by 5.17 just show closely MRA relate with quadratic mirror filtering.

The condition (5) in definition of MRA seems odd. In fact, it is the key to derive the expression form of wavelet function. We will skip all the technique part to show how we can find the mother wavelet function, starting from the knowledge of orthonormal basis $\varphi(\cdot - k)$. Interested reader can refer to Daubechies[1992], Chapter 5..

The following lemma, as a summary, directly gives the relation between mother wavelet function and scaling function.

Lemma 30 *Suppose $\varphi(\cdot)$ is a scaling function for an MRA $\{V_j\}_{j\in Z}$, and m_0 is the associated low-pass filter, which satisfy*

$$\hat{\varphi}(2\xi) = \hat{\varphi}(\xi)m_0(\xi)$$

then a function $\psi \in W_0 = V_1 \cap V_0^{\perp}$ is an orthonormal wavelet for $L^2(R)$ if and only if

$$\hat{\psi}(2\xi) = e^{i\xi}v(2\xi)\left[m_0(\xi+\pi)\right]^* \hat{\varphi}(\xi)$$

a.e. on R, for some $2\pi-$periodic measurable function v s.t.

$$|v(\xi)| = 1 \qquad\qquad a.e.\ on\ R.$$

(see Hernandez & Weiss P57 for Proof)

The above lemma formalizes the way to find the wavelet basis. In practice, the following general procedure is always adopted:

(see Daubechies[1992] P145)

(1) We always start from scaling function by choosing the function satisfying

1. (a) φ and $\hat{\varphi}$ have reasonable decay;

   (b) the following two conditions is satisfied:

      i. $\varphi(x) = \sum_{k\in Z} c_k \varphi(2x+k)$ where $\sum_k |c_k|^2 < \infty$ and

      ii. $0 < A \leq \sum_{k\in Z} |\hat{\varphi}(\xi + 2k\pi)|^2 \leq B < \infty$

   (c) $\hat{\varphi}(0) \neq 0$

The purpose of imposing the above condition (a) (b) (c) is obvious. Condition (a) assures the time and frequency localization of scaling function. Condition (i) of (b) is used to implement the condition (1) and (2) in the definition of MRA. Condition (ii) of (b) makes it possible to deploy the orthonormlization trick. Condition (c) is guarantee $l(\bigcup_{j\in Z} V_j) = L^2(R)$. Finally, with these condition, sequence $\{V_j\}$, where V_j is spanned by $\varphi(2^j \cdot +k)$ constitute a multiresolution analysis.

(2) If necessary, we need perform the 'orthonormalization trick'

$$\hat{\varphi}^{\#}(\xi) = \hat{\varphi}(\xi) \left[\sum_{k\in Z} |\hat{\varphi}(\xi + 2k\pi)|^2 \right]^{-1/2} \tag{5.18}$$

in order to make sure $\sum_{k\in Z} \left|\hat{\varphi}^{\#}(\xi + 2k\pi)\right|^2 = 1$[15].

(3) Finally, use the above lemma, we get

$$\hat{\psi}(\xi) = e^{i\xi/2} \left[m_0^{\#}(\xi/2 + \pi) \right]^* \hat{\varphi}^{\#}(\xi/2) \tag{5.19}$$

with

$$m_0^{\#}(\xi) = m_0(\xi) \left[\sum_{k\in Z} |\hat{\varphi}(\xi + 2k\pi)|^2 \right]^{1/2} \left[\sum_{k\in Z} |\hat{\varphi}(2\xi + 2k\pi)|^2 \right]^{-1/2}$$

[15] This equation directly lead to the conclusion that $\varphi^{\#}_{j,k}$ is orthonormal set.

or equivalently

$$\psi(x) = \sum_n (-1)^n h^{\#}_{-n+1} \varphi^{\#}(x-n), \tag{5.20}$$

with

$$m_0^{\#}(\xi) = \frac{1}{\sqrt{2}} \sum_n h_n^{\#} e^{-in\xi}$$

REFERENCES

Andersen, T.G. , Bollerslev, T, Diebold F.X. and Labys, P. ,2003, "Modeling and Forecasting Realized Volatility," Econometrica, 71, 579-625.

Antoniadis, A.; G. Oppenheim, Eds., 1995, Wavelets and statistics, Lecture Notes in Statistics 103, Springer Verlag.

Battle, 1987, "A block spin construction of ondelettes; Part I: Lemarie functions", Comm.Math.Phys., 110, 601-615.

Bierens, H.J., 1988, "ARMA Memory Index Modeling of Economic Time Series (with discussion)," Econometric Theory, 4, 35-59.

Bierens, H.J., 1990, "Model-Free Asymptotically Best Forecasting of Stationary Economic Time Series," Econometric Theory, 6, 348-383.

Birchenhall, Chris R. Jessen, Hans, Osborn Denise R. and Simpson, Paul, 1999, "Predicting U.S. Business Cycle Regimes," Journal of Business, 17, 313-323.

Bollerslev, T, Chou, R. Y. and Kroner, K. F., 1992, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," Journal of Econometrics, 52, 5-59.

Brier, G. W., 1950, "Verification of forecasts expressed in terms of Probability," Monthly Weather Review, 75, 1-3.

Burns, Arthur F. and Mitchell, Wesley C., 1946, "Measuring Business Cycle," New York, National Bureau of Economic Research.

Camacho, M. and Perez-Quiros, G., 2002, "This is What the Leading Indicators Lead," 17, 61-80.

Campbell, John Y, 1987, "Stock Returns and the Term Structure," Journal of Financial Economics, 18, 373-399.

Chui, C.K., 1992, "An introduction to Wavelets," Academic Press.

Cootner, Paul H., 1964, "The Random Character of Stock Market Prices," Cambridge, Mass, The M.I.T.Press.

Daubechies, 1992, " Ten Lectures on Wavelets," CBS-NSF Regional Conferences in Applied Mathematics, 61 SIAM.

De Bondt,Werner F.M. and Thaler,Richard, 1985, "Does the Stock Market Overreact?" Journal of Finance, 40, 793-805.

De Bondt,Werner F.M. and Thaler,Richard, 1987, “Further Evidence on Investor Overreaction and Stock Market Seasonality,” Journal of Finance, 42, 557-582.

Diebold F. X. and Rudebush G. D., 1989, “Scoring the Leading Indicators,” Journal of Business, 62, 369-391.

Donoho, D.L.; I.M. Johnstone, G. Kerkyacharian, D. Picard, 1996, “Density estimation by wavelet thesholding,” Annals of Stat., 24, 508-539.

Donoho, D.L., I.M. Johnstone, G. Kerkyacharian and D. Picard, 1995, “Wavelet shrinkage: asymptopia,” Jour. Roy. Stat. Soc., series B, 57, 301-369.

Dueker, Michael, 2002, “Regime-Dependent Recession Forecast Forecasts and the 2001 Recession,” Federal Reserve Bank of St Louis Review.

Estrella, A, and Mishkin F.S. , 1998, “Predicting US Recessions: Financial Variables as Leading Indicators,” The review of Economics and Statistics, 1, 45-61.

Fama, Eugene F. and French, Kenneth R., 1988, “Permanent and Temporary Components of Stock Market Prices,” Journal of Political Economy, 96, 246-273.

Filardo, Andrew J. , 1994, “Business-Cycle Phases and Their Transitional Dynamics,” Journal of Business & Economic Statistics, 12, 299-309.

Friedman, B.M. and Kuttner, K.N. ,1998, “Indicator Properties of the Paper-Bill Spread: Lessons from Recent Experience,” Review of Economics and Statistics, 80, 34-44.

Gabor, D, 1946, “Theory of Communication,” Journal of Institute Electric Engineering, London 93(III) 429-457.

Gallant, A.R., and D.W. Nychka, 1987, “Seminonparametric Maximum Likelihood Estimation,” Econometrica, 55, 363-390.

Gallant, A.R., and G. Tauchen, 1989, “Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications”, Econometrica, 57, 1091-1120.

George Bachman, Lawrence Narici, Edward Beckenstein, 2000, “Fourier and wavelet analysis,” New York, Springer.

Gordon, Robert J. Eds, 1986, “The American business cycle : continuity and change,” University of Chicago Press.

Grenander, Ulf. , 1981, “Abstract Inference”, Wiley.

Haar,A., 1910, "Zur theorie der orthogonalen funktionen systems," Math.Ann., 69, 331-371.

Hamilton, J & Jorda, O , 2002, "A Model for the Federal Funds Rate Target," Journal of Political Economy, 110, 1135-1167.

Hamilton, J., 1989, "A New Approach To The Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica, 57, 357-384.

Hamilton, J.D. and Perez-Quiros, G., 1996, "What Do Leading Indicators Lead," Journal of Business 69, 27-49.

Hamilton, James, 1994, "Time Series Analysis," Princeton Univ. Press.

Hansen, Peter R., 2001, "Test for Superior Predictive Ability," working paper, Brown University, Economics Department.

Harvey, Campbell, 1988, "The Real Term Structure and Consumption Growth," Journal of Financial Economics, 22, 305-333.

Harvey, Campbell and Akhtar Siddique, 2000, "Conditional Skewness in Asset Pricing Tests," Journal of Finance, 55, 1263-1295.

Hernandez and Weiss, 1996, A first course on Wavelets, CRC press.

Hirshleifer, David, 2001, "Investor Psychology and Asset Pricing," Journal of Finance, 56, 1533-1598.

Hymans, S.H., 1973, "On the Use of Leading Indicators to Predict Cyclical Turning Points", Brookings Papers on economic Activity, 2, 339-375.

Jagadeesh,Narasimhan and Titman,Sheridan, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," Journal of Finance, 48, 65-91.

Jagadeesh,Narasimhan, 1990, "Evidence of Predictable Behavior of Security Returns," Journal of Finance, 45, 881-898.

Jensen, Michael C., 1969, "Risk, the pricing of Capital Assets, and the Evaluation of Investment Portfolios," Journal of Business, 42, 167-247.

Kim C-J and Nelson C. R., 1998, "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on Dynamic Factor Model with Regime Switching," The Review of Economics and Statistics, 80, 188-201.

Kling, John L., 1987, "Predicting the Turning Points of Business and Economic Time Series," Journal of Business, 60, 201-238.

Koskinen, L and Oller, L, 2000, “A Classifying Procedure for Signaling Turning Point,” working paper, Stockholm School of Economics.

Lahiri, K. and Wang, J.G., 1994, “Predicting Cyclical Turning Points with a leading Index in a Markov Switching Model,” Journal of Forecasting, 13, 245-263.

Lahiri, Kajal and Moore, Geoffrey H., 1991, Leading Economic Indicators : New Approaches and Forecasting Records, Cambridge University Press.

Lakonishok, Josef and Shleifer, Andrei and Vishny,Robert W., 1994, “Contrarian Investment, Extrapolation, and Risk,” Journal of Finance, 49, 1541-1578.

Layton, A.P., 1996, “Dating and Predicting Phase Changes in the U.S. Business Cycle,” International Journal of Forecasting 12, 417-428.

Lehmann, Bruce N., 1990, “Fads, Martingales, and Market Efficiency,” Quarterly Journal of Economics, 105, 1-28.

Lemarie-Rieusset P.G , 1988, “Ondelettes a Localisation Exponentiells,” J.Math.Pure et Appl., 67, 227-236.

Lewellen, Jonathan, 2002, “Momentum and Autocorrelation in Stock Returns”, Review of Financial Studies, 15, 533-564.

Lo, Andrew W. Eds, 1997, “Market Efficiency : Stock Market Behaviour in Theory and Practice”, Edward Elgar Publishing.

Mallat, S., 1989, “Multiresolution approximations and wavelet orthonormal bases”, Trans.of Amer.Math.Soc., 315, 69-87.

McQueen,G. and Thorley, S , 1991, “Are Stock Returns Predictable ? A test using Markov Chains,” Journal of Finance, 46, 239-263.

Merton, R.C, 1973, “An Intertemporal capital asset pricing model,” Econometrica, 46, 867-887.

Meyer, Y, 1985, “Principe Dincertitude , Bases Hlbertiennes Et Algebras Doperateurs,” Seminaire Bourbaki, 1985-1986, no 662.

Meyer, Y, 1993, “Wavelets : Algorithms and Applications,” Society for Industrial and Applied Mathematics.

Neftci , N Salih, 1982, “Optimal Prediction of Cyclical Downturns,” Journal of Economic Dynamics and Control, 4, 225-241.

Niemira, M.P. & Klein , Philip A, 1994, "Forecasting Financial and Economic Cycles," John Wiley & Sons, Inc.

Percival, Donald B., 2000, "Wavelet methods for time series analysis," Cambridge University Press.

Politis, D. , Romano, J., 1994, "The Stationary Bootstrap," Journal of the American Statistical Association 89, 1303-1313.

Shefrin, Hersh, 2000, "Beyond Greed and Fear : Understanding Behavioral Finance and the Psychology of Investing," Harvard Business School Press.

Shleifer, Andrei, 2000, "Inefficient Markets : An Introduction to Behavioral Finance," Oxford University Press.

Stock, J. and Watson, M., 1993, "A Procedure for Predicting Recessions with Leading Indicator: Econometric Issues and Recent Performance." In : Stock and Watson, M.(Eds) Business Cycles, Indicators, and Forecasting. The University of Chicago Press, Chicago.

Stromberg, J.O , 1981, "A Modified Franklin System and Higher Order Spline Systems on As Unconditional Basis for Hardy Spaces," in Conference in honor of A.Zygmund (W.Beckner, Ed.), V.II, Wasdsword, 475-493.

Stuart Geman & Chii-Ruey Hwang, 1982, "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," The Annals of Statistics, 10, 401-414.

Sullivan, R., White, H. and Golomb, B., 2001, "Dangers of Data Mining: The Case of Calendar Effects in Stock Returns," Journal of Econometrics, 105, 249-286.

Tay , Anthony & Wallis, Kenneth F. , 2000, . "Density Forecasting: A Survey," Econometric Society World Congress 2000 Contributed Papers 0370.

Tchamitchian, PH, 1987, "Biorthogonalite et Theorie Des Operateurs," Rev Math. Iberoamer., 3 163-189.

Walter, Gilbert G., 2001, "Wavelets and other orthogonal systems," 2nd edition, Chapman & Hall/CRC.

Wayne Ferson and Campbell R. Harvey, 1991, "The Variation of Economic Risk Premiums," Journal of Political Economy, 99, 385-415.

Wecker, William F. ,1979, "Predicting the Turning Points of a Time Series," Journal of Business, 52, 35-50.

White, H., 1994, "Estimation, Inference, and Specification Analysis, " New York: Cambridge University Press.

VITA

Guang Guo was born in Shanghai, People's Republic of China, on November 13, 1973. He graduated from the High School affiliated with the Shanghai Normal School in Shanghai in June 1992. He received the B.S. degree, with departmental honors, in International Finance and Banking from Xiamen University in Fujian, P.R.C., in June 1996. From July 1996 through May 1998, he was employed as an investment consultant by Bank of Communications in Shanghai, P.R.C, evaluating and access the investment risk of commercial loan that involve foreign exchange. He enrolled in the Ph.D. program in Economics at the Pennsylvania State University in August 1998. Guang is married to Xiaofeng Wang and they have a son Max.