Prediction in Financial Market : The State Space Approach

By Hiroyuki Moriya*

This paper reviews statistical prediction models for financial time series. The classical statistical prediction theory developed by Wiener and Kolmogorov is applied to stationary stochastic processes. This type of model, an ARIMA model, is practical and thus widely used for financial time series analysis. However, most financial time series are nonstationary. State space models, extensively used in the field of control engineering, can create an approach for a wide range of complex financial time series. Unfortunately, this approach is not actively used in the financial industry. The paper explains the basic properties of state space models, structure modeling, and a recursive algorithm called the Kalman filter for novice researchers and practitioners.

1 Introduction

Predictions in the financial industry are divided into several categories. One way to forecast market movements is by seeing analogies in historical price movements or patterns: observing similar market movements in, for example, previously observed price patterns leads to the forecast that subsequent market movements will be similar to those previously observed. The"forecast" is the simple assumption that previous events may repeat in the future.

This predictive statement, based on the assumption of the repetition or continuation of patterns arbitrarily or without motivation, is strongly subjective. However, as such predictive statements are based on stylized facts, these methods may be classified as a scientific, qualitative prediction.

Price predictions are often made by computing a moving average of a number of historical prices. In practice, the number of past prices used is often not optimized but follows a"rule of thumb" which may optimize moving averages. Holt

*This work is based on the discussions with Professor Hiroshi Tsuda, Doshisha University, Adjunct Professor Yoshiaki Kumagai, Waseda University, and a comment by Professor Yuji Aruka, Chuo University. The basic idea of this work owes a debt to Andrew C. Harvey's extensive research works, especially his book "Forecasting, structural time series models and the Kalman filter"

(1957) and Winters (1960) introduced an updating method of an exponentially weighted moving average for forcasting a local linear trend.

Then there is prediction based on a hypothesis or a model and related to stationary stochastic processes. The prediction is a linear function of past values with the minimization of the mean-square prediction error, as developed by Wiener (1949) and Kolmogorov (1941) and explained by Whittle (1963). This methodology developed into as the Box-Jenkins ARIMA model (1970) that is the main analytical method currently in use for financial time series analysis in the industry. It assumes that a series is reduced to stationarity by differencing.

The state space method is completely different from the Box-Jenkins approach. It has submodels designed to capture movements of observations consistent with its properties. These submodels are components that are not observable and have a stochastic form. Each unobserved component might be described as a trend, a season, a cycle or a disturbance term, and is called a state of the system. It is difficult to identify all factors or components in order to generate the observations. This statistical approach of decomposing a time series model into several components is structure modeling. A simple starting point of a state space model is that time series data are decomposed into a trend and a disturbance term.

If a prediction function does not have random variables but has definite predictive values, that prediction is known as deterministic. On the other hand, when a prediction function is described as "a random walk," as for stock price movements, then it contains a stochastic process: the predictive function is not deterministic but is a conditional probability distribution of future values based on a knowledge of historical prices.

A stochastic process is useful when one is interested in an average or in typical behavior; for example, an analyst researches gold prices by calling major buyers and sellers in the relevant industries, and then determining not player-specific but average demand and supply information. In another example, analysts often observe complicated and unexplained variations in a financial market. In such cases, the model has a residual term that represents the unexpected variation of predictive values. The residual is thus assumed to follow a stochastic process. A stochastic process often makes a deterministic function evolve over time, and the evolution of the stochastic term will determine the properties of the time series model.

In a state space model, when a new observation comes in, the Kalman filter computes an optimal estimator for the system. Proposed by Kalman (1960), it is a recursive procedure used extensively for certain engineering applications. It is not actively applied in the financial industry, however the approach has been popularized gradually due to its usefulness and to recent developments in computational power.

To predict is to make a statement of what or how an object will be in the future. At times, prediction is a major objective for time series modeling; however, it is rarely an end in itself.

2 The state space approach

In a statistical model, time series are assumed to follow certain probability distributions. Such a model captures major characteristics of the data generation process. For example, a random walk model is written as $y_t = y_{t-1} + \epsilon_t$. In this model, the disturbance term ϵ_t follows the normal distribution with mean zero and variance σ_ϵ^2. The model has two parameters: the mean and variance. These two values are assumed to be fixed and unknown in the classical analysis. This model is written as:

$$y_t = y_1 + \sum_{j=1}^{t} \epsilon_j,$$

where $dy = y_t - y_{t-1} = \epsilon_t$ is assumed to be stationary, $dy \sim N(0, \sigma_\epsilon)$, but y_t itself is nonstationary.

A random walk model has a stochastic trend. Applying the statistical theory, the level of the stochastic trend is obtained as follows: at time $t = 1$, y_1 follows the normal distribution with mean a_1 and variance Q_1, that is, $y_1 \sim N(a_1, Q_1)$. When $t = 2$, y_2 follows the normal distribution with a_2 and Q_2. This data generation process is expressed as $\{y_1, y_2\}$, and is pass-dependent. Its probability is given by

$$p(y_1, y_2) = p(y_1)p(y_2 \mid y_1),$$

where $p(y_2 \mid y_1)$ represents the conditional probability of y_1 followed by y_2. As time passes new observations are brought in, then a sequence of y_t is denoted as $Y_t = \{y_1, y_2, \ldots, y_t\}$. The probability of Y_t is written as

$$\begin{aligned} p(y_1, \ldots, y_t) &= p(Y_t) = p(y_1) \prod_{n=2}^{t} p(y_n \mid Y_{n-1}) \\ &= \prod_{n=1}^{t} p(y_n \mid Y_{n-1}), \end{aligned} \tag{1}$$

where $p(y_1 \mid Y_0) = p(y_1)$. y_t follows the normal distribution of mean a_t and variance Q_t. The values of a_t and Q_t may vary according to a given sample Y_t, as that sample represents the level of the stochastic trend with a given Y_t.

Now we rewrite a random walk model in state-space form to describe the property of a state-space model that consists of a measurement equation and a transition (state) equation;

$$y_t = \alpha_t + \epsilon_t \tag{2}$$

$$\alpha_{t+1} = \alpha_t + \eta_t., \tag{3}$$

where y_t is observed, and α_t is unobserved. Equation (2) represents the relationship between observed data y_t and unobserved data α_t plus noise ϵ, and is known as the measurement equation. It provides the link between the sample and the unobserved variable.

In equation (3), the dynamic of the unobserved variable is described as the form of a first-order difference equation. α_t is modeled stochastically as a function of disturbance. It represents the stochastic trend or the level of trend conditional on past observations. Unobserved variables are a "state of the system". The variances of ϵ_t and η_t are the parameters of the system and the determinants of its properties. They are known as hyperparameters, generally denoted by Ψ. A set of equations (2) and (3) is known as a local level model.

The transition equation is a random walk; therefore the measurement equation is also. Since a random walk is nonstationary, equations (2) and (3) are nonstationary. Thus, if y_t and α_t follow the normal distribution, the parameters of these distributions change over time. Rewriting the local level model:

$$y_t = \alpha_1 + \sum_{n=1}^{t} \eta_n + \epsilon_t.$$

The most prominent feature of the state-space model is its flexibility. If the model has several regression coefficients, they can vary over time to be consistent with the characteristics of observations for each component of the model. In the classical linear regression, we assume that parameters are fixed but unknown. They do not change over time. In the state-space model we can assume that the parameters are fixed but unknown, as in the classical regression analysis, but we can also assume that the parameters are changed over time. It can be a time-invariant as well as a time-varying model.

A set of equations (2) and (3) can describe the data generation process of y_t with emphasis on the prediction's accuracy. The recursive techniques known as the Kalman filter may provide optimal predictors. Unobserved random variable α_t fluctuates around the mean with the range given by the variance. Since α_t is an unobserved variable, ϵ_t and η_t are unobserved as well. However, computing the expected value of α_t produces the following values :

$$\begin{aligned} \hat{\epsilon}_t &= y_t - \hat{\alpha}_t, \\ \hat{\eta}_t &= \hat{\alpha}_{t+1} - \hat{\alpha}_t, \end{aligned}$$

where $\hat{\alpha}_t$ is the expected value of α_t, and $\hat{\epsilon}$ and $\hat{\eta}$ are known as disturbance smoothing and state smoothing, respectively. $\hat{\eta}_t$ is useful to construct confidence intervals for α_t. We can detect outliers and structural breaks by analyzing $\hat{\epsilon}_t$ and $\hat{\eta}_t$.

2.1 Assumptions of the model

The set of equations (2) and (3) consists of the local level model. It is the simplest state space model decomposed into a trend and a disturbance term. In this model, it is assumed that all variables are normally distributed so that any combinations of joint variables are normally distributed. It is also assumed that ϵ_t and η_t have constant variances σ_ϵ^2 and σ_η^2 and that ϵ_t and η_t are mutually independent and independent of α_t.

$$\epsilon_t \sim N(0, \sigma_\epsilon^2),$$

$$\eta_t \sim N(0, \sigma_\eta^2).$$

Initially, we know that α_1 follows the normal distribution with mean a_1 and variance P_1.

$$\alpha_1 \sim N(a_1, P_1).$$

At time $t-1$ and a given Y_{t-1}, α_t is assumed to follow the normal distribution with mean a_t and variance P_t:

$$\alpha_t \sim N(a_t, P_t).$$

The subscript represents the change of parameters. As the property of a random walk process, a_t and P_t vary as time passes. We can express

$$\begin{aligned} a_t &= E(\alpha_t \mid Y_{t-1}), \\ P_t &= \mathrm{Var}(\alpha_t \mid Y_{t-1}). \end{aligned}$$

We define $v_t = y_t - a_t$ and $V_t = \sigma_{v_t}^2$. v_t is called an innovation because it explains unpredictable value that cannot be obtained from the function of past observations. V_t is a measure of the size of innovative impact at time t. Replacing equation (2) for y_t [1], we obtain the expected value and the variance of v_t:

$$\begin{aligned} E(v_t \mid Y_{t-1}) &= E(y_t - a_t \mid Y_{t-1}) \\ &= E(\alpha_t + \epsilon_t - a_t \mid Y_{t-1}) \\ &= E(\alpha_t \mid Y_{t-1}) + E(\epsilon_t \mid Y_{t-1}) - E(a_t \mid Y_{t-1}) \\ &= a_t - a_t = 0, \\ \mathrm{Var}(v_t \mid Y_{t-1}) &= V_t = \mathrm{Var}(y_t - a_t) \\ &= \mathrm{Var}(y_t) - \mathrm{Var}(a_t) - 2\mathrm{Cov}(y_t, a_t) \\ &= \mathrm{Var}(y_t), \\ \mathrm{Var}(v_t \mid Y_{t-1}) &= V_t = \mathrm{Var}(y_t - a_t) \\ &= \mathrm{Var}(\alpha_t + \epsilon_t - a_t) \\ &= \mathrm{Var}(\alpha_t \mid Y_{t-1}) + \sigma_\epsilon^2 \\ &= P_t + \sigma_\epsilon^2, \end{aligned}$$

and

$$\begin{aligned} \mathrm{Cov}(v_t, y_i) &= E(v_t y_i) - E(v_t)E(y_i) \\ &= E(v_t y_i) \\ &= E[E(v_t \mid Y_{t-1}) y_i] = 0, \end{aligned}$$

[1] We use the following properties of expectation and variance:

$$\begin{aligned} E(X+Y) &= E(X) + E(Y), \\ \mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y). \end{aligned}$$

then

$$v_t \sim NID(0, V_t).$$

Therefore we assume that the innovation follows the normal distribution with mean zero and variance V_t.

2.2 The Kalman filter

Initially, it is assumed that α_1 follows the normal distribution with mean a_1 and variance P_1. When the random variable y_1 is observed at time $t = 1$, it is compared to the estimate a_1. Based on y_1, a_1 and P_1 are updated and the updated values are used to compute new prediction values of a_2 and P_2.

At time $t = 2$, there is a new observation y_2. Then, a_2 and P_2 are updated and used to calculate a_3 and P_3. In this way, the local level of α_t is updated based on actual observations.

2.2.1 Updating the previous prediction

When we update a_t and P_t from newly observed y_t, we assume that v_t and $E(\alpha_t \mid Y_t)$ have a linear relationship with the intercept a_t.

$$a_t^{update} = a_t + \beta_a v_t = a_t + \beta_a(y_t - a_t),$$

where $a_t^{update} = E(\alpha_t \mid Y_t)$ and β_a is a slope coefficient. This is a reasonable assumption because of $E(v_t) = 0$. The expected value of α_t at a given Y_t is expressed as

$$\begin{aligned} a_t^{update} &= E(\alpha_t \mid Y_{t-1}, v_t) \\ &= E(\alpha_t \mid Y_{t-1}) + \frac{\mathrm{Cov}(\alpha_t, v_t)}{\sigma_{v_t}^2} v_t, \end{aligned} \tag{4}$$

where

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_t) &= E(\alpha_t v_t) - E(\alpha_t)E(v_t) \\ &= E(\alpha_t\alpha_t) + E(\alpha_t\epsilon_t) \\ &\quad -E(\alpha_t a_t) - E(\alpha_t)E(\alpha_t) - E(\alpha_t)E(\epsilon_t) + E(\alpha_t)E(a_t) \\ &= E(\alpha_t\alpha_t) - E(\alpha_t)E(\alpha_t) \\ &= \mathrm{Var}(\alpha_t \mid Y_{t-1}) \\ &= P_t. \end{aligned}$$

Now equation (4) is written as

$$a_t^{update} = a_t + K_t^x v_t, \tag{5}$$

where $K_t^x = P_t/V_t$ and is called the Kalman gain. This is the regression coefficient of α_t on v_t. The update from $E(\alpha_t \mid Y_{t-1})$ based on y_t is a_t^{update}.

We can develop the same form of the variance of α_t.

$$P_t^{update} = P_t + \beta_A V_t,$$

where $P_t^{update} = \mathrm{Var}(\alpha_t \mid Y_t)$ and β_A is a slope coefficient.

$$\begin{aligned} P_t^{update} &= \mathrm{Var}(\alpha_t \mid Y_{t-1}, v_t) \\ &= \mathrm{Var}(\alpha_t \mid Y_{t-1}) - \mathrm{Cov}(\alpha_t, v_t)^2 \mathrm{Var}(v_t)^{-1} \\ &= P_t - P_t^2/V_t. \end{aligned} \tag{6}$$

Equation (5) and (6) are known as updating equations.

2.2.2 New prediction

The optimal predictor one step ahead of α_t is the expected value of α_{t+1} and its variance is the corresponding variance. To estimate a_{t+1} and P_{t+1} as

$$\begin{aligned} a_{t+1} &= E(\alpha_{t+1} \mid Y_t) \\ &= E(\alpha_t + \eta_t \mid Y_t) \\ &= E(\alpha_t \mid Y_t) + E(\eta_t \mid Y_t), \\ P_{t+1} &= \mathrm{Var}(\alpha_{t+1} \mid Y_t) \\ &= \mathrm{Var}(\alpha_t + \eta_t \mid Y_t) \\ &= \mathrm{Var}(\alpha_t \mid Y_t) + \mathrm{Var}(\eta_t \mid Y_t). \end{aligned}$$

we obtain

$$\begin{aligned} a_{t+1} &= a_t^{update} = E(\alpha_t \mid Y_t), \\ P_{t+1} &= P_t^{update} + \sigma_\eta^2 = \mathrm{Var}(\alpha_t \mid Y_t) + \sigma_\eta^2. \end{aligned}$$

These two equations are called prediction equations.

Finally we get a set of formulae:

$$\begin{aligned} v_t &= y_t - a_t, \\ V_t &= P_t + \sigma_\epsilon^2, \\ K_t^x &= P_t/V_t, \\ a_{t+1} &= a_t + K_t v_t, \\ P_{t+1} &= P_t(1 - K_t) + \sigma_\eta^2. \end{aligned}$$

This set of equation makes up the Kalman filter. Please see Harvey(1989) and Durbin and Koopman(2001) for more detailed discussion about the local level model.

2.3 Initialization

The set of relations that constitutes the Kalman filter is known as a recursion. In a recursive process, the initial values of some variables must be assumed

or chosen. For example, in the local level model, determining the value of a_1 and P_1 for the unknown distribution of α_1 specifies a diffuse prior distribution. This process is purely subjective. There are two methods to determine the parameters of a_1 and P_1.

We fix a_1 at an arbitrary value and $P_1 \to \infty$.

$$\begin{aligned} a_2 &= a_1 + K_1^x v_1 \\ &= a_1 + \frac{P_1}{P_1 + \sigma_\epsilon^2}(y_1 - a_1) \\ &= y_1, \\ P_2 &= P_1(1 - K_1^x) + \sigma_\eta^2 \\ &= \frac{P_1}{P_1 + \sigma_\epsilon^2}\sigma_\epsilon^2 + \sigma_\eta^2 \\ &= \sigma_\epsilon^2 + \sigma_\eta^2. \end{aligned}$$

This process is known as diffuse initialization and this filter is called the diffuse Kalman filter.

Maximum likelihood may be defined as a statistical method for estimating true parameters from the given data. We can estimate the parameters using maximum likelihood from the first observation y_1. The estimates of the parameters maximize the probability of obtaining the observations. In this case, y_1 is the only sample. The Kalman filter is initialized by taking $a_1 = \alpha_1 = y_1$ and $P_1 = \mathrm{Var}(\alpha_1) = \mathrm{Var}(y_1)$. Here, α_1 and $\mathrm{Var}(\alpha_1)$ are maximum likelihood estimates.

If α_1 is constant, then α_1 is fixed and $y_1 \sim NID(\alpha_1, \sigma_\epsilon^2)$, and then $\alpha_1 = y_1$ and $\mathrm{Var}(\alpha_1) = \sigma_\epsilon^2$.

If the initial values are above the true values, the model may not diverge over time. However this will make the filter less sensitive.

2.4 Estimation of hyperparameters

From equation (1), the probability of $Y_t = \{y_1, y_2, \ldots, y_{t-1}, y_t\}$ is given by :

$$p(y_1, \ldots, y_t) = \prod_{n=1}^{t} p(y_n \mid Y_{n-1}).$$

Then, the loglikelihood is

$$\log L = \log p(Y_t) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{n=1}^{t}(\log V_t + \frac{v^2}{V_t}).$$

Maximizing the loglikelihood with respect to the parameters σ_ϵ^2 and σ_η^2, we use an EM algorithm to find their true values. The EM algorithm is a numerical maximization method that consists of an expectation step (E) and a maximization step (M). The algorithm evaluates and then maximizes the conditional

expectation with respect to parameters σ_ϵ^2 and σ_η^2, as developed by Shumway and Stoffer (1982), Watson and Engle (1983) and Koopman (1993).

Putting $q = \sigma_\eta^2/\sigma_\epsilon^2$, the local level model may be reformulated as:

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2), \\ \alpha_{t+1} &= \alpha_t + q\epsilon_t, \quad \eta_t \sim NID(0, q\sigma_\epsilon^2), \\ y_t &\sim N(a_t, V_t), \\ \alpha_t &\sim N(a_t, P_t), \end{aligned}$$

where q is known as a signal-to-noise ratio. Restating a set of formulae for the Kalman filter:

$$\begin{aligned} v_t &= y_t - a_t, \\ V_t &= P_t + \sigma_\epsilon^2, \\ K_t^x &= P_t/V_t = P_t^*/V_t^*, \\ a_{t+1} &= a_t + K_t v_t, \\ P_{t+1} &= P_t(1 - K_t) + q\sigma_\epsilon^2, \\ V_t^* &= V_t/\sigma_\epsilon^2 = P_t^* + 1, \\ P_{t+1}^* &= P_t^*(1 - K_t^x) + q. \end{aligned}$$

As y_t follows $N(a_t, V_t)$, we can maximize the loglikelihood function of

$$\begin{aligned} \log L^* &= -\frac{1}{2}\log 2\pi - \frac{n-1}{2} - \frac{n-1}{2}\log\hat{\sigma}_\epsilon^2 - \frac{1}{2}\sum_{t=2}^{n}\log V_t^* \\ &= -\frac{1}{2}\log 2\pi - \frac{n-1}{2} - \frac{n-1}{2}\log\hat{\sigma}_\epsilon^2 - \frac{1}{2}\sum_{t=2}^{n}\log[P_{t-1}^*(1 - K_{t+1}) + q], \end{aligned}$$

where

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-1}\sum_{t=2}^{n}\frac{v_t^2}{V_t^*}.$$

$\log L^*$ is referred to as the concentrated diffuse loglikelihood, and is maximized with respect to q. It is used in the Broyden-Fletcher-Goldfarb-Shannon (BFGS) method to optimize $\log L^*$, and is a type of Newtonian method of optimization. The detailed discussion can be found in Fletcher (1987).

2.5 Steady state of the local level model

Does the Kalman filter converge to a steady state as $n \to \infty$? It means that the system parameters converge to certain values, for example

- $V_t \to P_t + \sigma_\epsilon^2$,
- $K_t \to P_t/V_t = P_t/(P_t + \sigma_\epsilon^2)$,

- $P_t \to P_t(1-K_t) + \sigma_\eta^2 = P_t\left[1 - P_t/(P_t+\sigma_\epsilon^2)\right] + \sigma_\eta^2$.

The variance of α_{t+1}, P_{t+1}, is given by

$$P_{t+1} = P_t[1 - P_t/(P_t+\sigma_\epsilon^2)] + \sigma_\eta^2.$$

If the system is in the steady state, it is said to be

$$\bar{P} = P_{t+1} = P_t,$$

where $\bar{P}$ is the steady state of the variance of α_t as $t \to \infty$. Then the formula above is restated as

$$\bar{P} \quad = \quad \bar{P}[1 - \bar{P}/(\bar{P}+\sigma_\epsilon^2)] + \sigma_\eta^2,$$

thus

$$(P^*)^2 - (P^*)q - q = 0,$$

where $P^* = \bar{P}/\sigma_\epsilon^2$, q is the signal-to-noise ratio. Its solution is

$$P^* = \left(q \pm \sqrt{q^2+4q}\right)/2 > 0.$$

Thus when $q > 0$, then the solution is positive.

Next we consider the extreme case of $\sigma_\epsilon^2 = 0$. That is, we assume that the values given by the local level model and the values of the observations are identical and without any errors; then

- $y_{t+1} = \alpha_{t+1}$,
- $V_t = P_t$,
- $K_t^x = P_t/V_t = 1$,
- $a_{t+1} = a_t + K_t^x v_t = a_t + y_t - a_t = y_t$, and
- $P_{t+1} = \sigma_\eta^2$.

The expected value y_{t+1} is today's value y_t. The variance of α_t stays constant as the value of σ_η^2.

Let us consider another extreme case in which $\sigma_\eta^2 = 0$: α_t becomes a deterministic process, and then

- $V_t = P_t + \sigma_\epsilon^2 \to 0$,
- $K_t^x = P_t/V_t \to 0$,
- $a_{t+1} = a_t + K_t^x v_t \to a_1$, and
- $P_{t+1} = P_t(1-K_t) + \sigma_\eta^2 \to 0$.

However, these two extreme examples are very rare events. In general,

$$\text{when} \quad \sigma_\epsilon^2 > \sigma_\eta^2, \quad \text{then} \quad K_t^x \to 0,$$

$$\text{when} \quad \sigma_\epsilon^2 \leq \sigma_\eta^2, \quad \text{then} \quad K_t^x \to 1.$$

3 The general linear Gaussian state space model

Analyzing the local level model demonstrates basic properties of state space form, leading to the basic formation of the state space model. The observations contain N elements and are decomposed into m components, applied to a multivariate time series. When N is equal to 1, it is a univariate model.

$$\begin{aligned} \mathbf{y_t} &= \mathbf{Z_t}\alpha_\mathbf{t} + \mathbf{d_t} + \epsilon_\mathbf{t}, \\ \alpha_\mathbf{t} &= \mathbf{T_t}\alpha_\mathbf{t-1} + \mathbf{c_t} + \mathbf{R_t} + \eta_\mathbf{t}, \end{aligned} \tag{7}$$

where $\mathbf{Z_t}, \mathbf{d_t}, \mathbf{T_t}, \mathbf{c_t}$, and $\mathbf{R_t}$ are system matrices. In the measurement equation, $\mathbf{Z_t}$ is an $N \times m$ matrix, $\mathbf{d_t}$ is an $N \times 1$ vector, and $\epsilon_\mathbf{t}$ is an $N \times 1$ vector of serially uncorrelated disturbances with mean = zero and covariance matrix $\mathbf{H_t}$. In the transition equation, $\mathbf{T_t}$ is an $m \times m$ matrix, $\mathbf{c_t}$ is an $m \times 1$ vector, $\mathbf{R_t}$ is an $m \times r$ matrix, and $\eta_\mathbf{t}$ is an $r \times 1$ vector of serially uncorrelated disturbances with mean = zero and covariance matrix $\mathbf{Q_t}$. In this system, the observations $\mathbf{y_t}$ are modeled as a linear combination of disturbances over time and the initial state vector, $\alpha_\mathbf{0}$. $\mathbf{d_t}$ is an explanatory variable that enables the model to include observable variables to explain some of the movement in the observations, $\mathbf{y_t}$. It might increase flexibility because the explanatory variable may take time-varying coefficients. The disturbances $\epsilon_\mathbf{t}$ and $\eta_\mathbf{t}$ and the initial state vector $\alpha_\mathbf{0}$ are assumed to be normally distributed. These disturbances are uncorrelated with each other and uncorrelated with the initial state over time. Note that the system matrices take subscript t, meaning that the model may take more than one state. Thus

$$\mathrm{cov}(\epsilon_\mathbf{t}, \eta_\mathbf{j}) = \mathbf{0} \quad \text{for all } j \text{ and } t,\ t \neq j,$$

where j denotes time. The parameters appearing in the system matrices $\mathbf{Z_t}$,$\mathbf{T_t}$, $\mathbf{R_t}$, $\mathbf{H_t}$ and $\mathbf{Q_t}$ determine the properties of the model. They are called the hyperparameters and are distinguished from the parameters appearing in the system matrices $\mathbf{d_t}$ and $\mathbf{c_t}$. The later parameters determine only the characteristics of their own states. This model is called the general linear Gaussian state space model.

In the general state space form, the Kalman filter computes the optimal estimator of the state vector as time passes, as previously described.

The hyperparameters appearing in the system matrices and initial values of $\mathbf{a_0}$ and $\mathbf{P_0}$ are fixed and unknown. We have to estimate these values before the system is applied.

The optimal estimators of $\alpha_\mathbf{t}$ at $t-1$ are denoted as

$$\begin{aligned} \mathbf{a_t} &= E(\alpha_\mathbf{t} \mid \mathbf{Y_{t-1}}), \\ \mathbf{P_t} &= \mathrm{Var}(\alpha_\mathbf{t} \mid \mathbf{Y_{t-1}}). \end{aligned}$$

$\mathbf{P_t}$ can be regarded as its unconditional error covariance matrix.

When new observations $\mathbf{y_t}$ are brought in, the predictive values of $\mathbf{a_t}$ and

$\mathbf{P_t}$ are updated by

$$\begin{aligned}\mathbf{a_t^{update}} &= E(\alpha_\mathbf{t} \mid \mathbf{Y_t}) = \mathbf{a_t} + \mathbf{P_t Z_t' F_t^{-1}}(\mathbf{y_t} - \mathbf{Z_t a_t} - \mathbf{d_t}),\\ \mathbf{P_t^{update}} &= \mathrm{Var}(\alpha_\mathbf{t} \mid \mathbf{Y_t}) = \mathbf{P_t} - \mathbf{P_t Z_t' F_t^{-1} P_t},\end{aligned}$$

where $\mathbf{V_t} = \mathbf{Z_t P_t Z_t'} + \mathbf{H_t}$. These two equations are called the updating equations.

Based on new estimators of $\mathbf{a_t^{update}}$ and $\mathbf{P_t^{update}}$, we make an optimal prediction for the next step:

$$\begin{aligned}\mathbf{a_{t+1}} &= E(\alpha_{\mathbf{t+1}} \mid \mathbf{Y_t}) = \mathbf{T_t a_t^{update}} + \mathbf{c_{t+1}},\\ \mathbf{P_{t+1}} &= \mathrm{Var}(\alpha_{\mathbf{t+1}} \mid \mathbf{Y_t}) = \mathbf{T_t P_t^{update} T_t'} + \mathbf{R_t Q_t R_t'}. \qquad (8)\end{aligned}$$

These prediction equations are known as the Kalman filter. After the completion of the recursive process of the Kalman filter, $\mathbf{a_t}$ is the optimal estimator of $\alpha_\mathbf{t}$, with the minimum mean square error(MMSE).

3.1 The time invariant framework

The local level model is a time-invariant model: the system matrices are fixed over time. In many applications, the system matrices are assumed to be non-stochastic. Even though time passes, the hyperparameters and the properties of the system remain constant. Under such an assumption, the system matrices are denoted as $\mathbf{Z}, \mathbf{d}, \mathbf{T}, \mathbf{c}$, and $\mathbf{R}$.

In a stationary process, the parameters stay constant so that the process's properties do not change over time. Many nonstationary time series data are modeled as time-invariant forms and are made stationary by differencing, decomposing into several components, or including explanatory variables in the model. The models in state space form classified as time-invariant are much broader than the stationary models in which stationarity is obtained simply by differencing.

The definition of $\alpha_\mathbf{t}$ may determine the stationarity of the model. The state system should contain all the relevant information, however the number of state variables or elements must be minimized, making it depend on the interpretation of observations, depth of experience and computational power at that time.

If the state system converges with the equilibrium condition, the system will be stable. Such a condition will be achieved depending on the initial value of $\alpha_\mathbf{0}$, and the properties of $\epsilon_\mathbf{t}$ and $\eta_\mathbf{t}$. However regardless of what value $\alpha_\mathbf{0}, \epsilon_\mathbf{t}$ and $\eta_\mathbf{t}$ take, the necessary and sufficient conditions for stability is

$$\mathbf{T} < \mathbf{1} \text{ for all components.}$$

Please see Harvey (1989,3.3).

3.2 Another form of trends

Defining a trend is a very important issue in applied work. Other kinds of trends can show how flexible the state space form is. These models are compared with the ARIMA and EWMA models in later sections.

3.2.1 The global trend

The global trend may be simply defined as a straight line fitting a function of time to all points in time series data. It is obtained by the linear regression method, and is a deterministic function of time. It is expressed as

$$y_t = g \cdot t + \epsilon_t,$$

where g is a regression coefficient and ϵ_t is a disturbance term. This is the most basic form of linear trend; the forecast extrapolates it into the future. A random walk with the global trend will be expressed in state space form as

$$\begin{aligned} y_t &= \alpha_t + g \cdot t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2), \\ \alpha_{t+1} &= \alpha_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \end{aligned}$$

thus we get

$$y_t = \alpha_1 + \sum_{n=1}^{t-1} \eta_n + g \cdot t + \epsilon_t.$$

The model contains both the local and global trends.

3.2.2 The local linear trend

A random walk model plus a time trend in state space form is expressed as

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2), \tag{9}$$

$$\alpha_{t+1} = \alpha_t + l_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \tag{10}$$

$$l_{t+1} = l_t + \lambda_t, \quad \lambda_t \sim NID(0, \sigma_\lambda^2). \tag{11}$$

This is a local linear trend model. From these three equations, it is rewritten as

$$\begin{aligned} y_t &= \alpha_0 + \sum_{n=0}^{t-1} \eta_n + \sum_{n=0}^{t-1} l_n + \epsilon_t \\ &= \alpha_0 + \sum_{n=0}^{t-1} \eta_n + l_0 \cdot t + \sum_{n=0}^{t-1} \sum_{m=0}^{n-1} \lambda_m + \epsilon_t \end{aligned}$$

where the coefficients or slope of the time trend is stochastic and varying over time.

If $\sigma_\lambda^2 = 0$, then the time trend becomes deterministic, with a fixed slope.

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2), \\ \alpha_{t+1} &= \alpha_t + l + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2). \end{aligned}$$

It is expressed as

$$y_t = \alpha_1 + \sum_{n=1}^{t-1} \eta_n + l \cdot t + \epsilon_t,$$

in which case it is seen as a random walk with drift, and the local trend might be identical to the global trend.

3.2.3 The higher-order polynomial trends

A local nonlinear trend is given by higher order polynomials. It is described as

$$y_t = l_0 + l_1 t + l_2 \frac{t^2}{2} +, \ldots, + l_n \frac{t^n}{n!} + \epsilon_t, \quad t = 1, \ldots, T.$$

This is a deterministic model. This kind of model can be formulated as a linear time series model in state space form. One of the simplest examples of this local trend model is constructed in state space form by setting:

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2), \\ \alpha_{t+1} &= \alpha_t + l_{0,t} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \\ l_{0,t+1} &= l_{0,t} + l_{1,t} + \psi_{0,t}, \quad \psi_{0,t} \sim NID(0, \sigma_{\psi_0}^2), \\ l_{1,t+1} &= l_{1,t} + \psi_{1,t}, \quad \psi_{1,t} \sim NID(0, \sigma_{\psi_1}^2). \end{aligned}$$

In this case, the coefficients l_0 and l_1 are stochastic. l_0 and l_1 are interpreted as the local trend and a local acceleration, respectively. This is a local quadratic trend model. y_t is expressed as

$$\begin{aligned} y_t &= \alpha_1 + \sum_{n=1}^{t-1} \eta_n + \sum_{n=0}^{t-1} l_{0,n} + \sum_{n=1}^{t-1} l_{1,n} \cdot t + \epsilon_t \\ &= \alpha_1 + \sum_{n=1}^{t-1} \eta_n + l_{0,1} \cdot t + + \sum_{n=0}^{t-1} \sum_{m=0}^{n-1} \psi_{0,m} \\ &\quad + l_{1,1} \cdot t^2 + \left(\sum_{n=1}^{t-1} \sum_{m=1}^{n-1} \psi_{1,m} \right) \cdot t + \epsilon_t. \end{aligned}$$

Consider the case of $\sigma_{\psi_0}^2 = 0$ and $\sigma_{\psi_1}^2 = 0$, implying l_0 and l_1 become constant and known as a random walk with drift plus a deterministic trend. If $l_1 = 0$, the model is known as a random walk with drift (which we have already seen). If $l_0 = 0$, the model is known as a random walk plus a deterministic trend. Please refer to the discussion about types of higher-order polynomial trends in Jacobs and Jones (1980).

So what is the trend? Harvey (1989, 6.1.1) clearly defines that in structural time series models, the trend indicates the long-term movements in the series, without seasonal and daily effects, that is the best estimator of the trend and tends towards the optimal predictor of the series as a whole. Symbolically it is expressed as

$$\lim_{i \to \infty} (\tilde{y}_{T+l|T} - m_{T+l|T}) = 0,$$

where $\tilde{y}_{T+l|T}$ is the optimized prediction of the series as a whole, and $m_{T+l|T}$ is the minimum mean square linear estimator(MMSLE) of the trend at time T.

3.3 Relationship to ARIMA and EWMA models

The structural time series model may have several disturbance terms. Since the model is linear, the state space model is combined to be a single equation with a single disturbance. This is known as a reduced form or a canonical form.

3.3.1 ARIMA models

In a financial time series, the lagged return or disturbance might be useful for describing the property of given data. It is modeled as an autoregressive-moving average process with order (p, q) that is given by

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p + \xi_1 + \theta_1 \xi_{t-1} + \cdots + \theta_q \xi_{t-1}.$$

It is denoted by ARMA(p, q).

A series might be stationary by differencing d times; it is said to be the integrated order of d. When a given sample is modeled to follow a stationary and an ARMA(p, q), after taking a difference d times, it is known as an autoregressive-integrated-moving average process with order (p, d, q) denoted as ARIMA(p, d, q). Please see Box and Jenkins (1970). A structural time series model is reduced to be an ARIMA(p, d, q) process, possibly with some restrictions on the parameter space. If there is no restriction in ARIMA models fitted to the given data, it is known as an unrestricted reduced form. Please refer to Nerlove (1979,pp.70-78) for a general algorithm .

3.3.2 Invertibility

An ARMA process is said to be invertible if

$$\xi_t = \sum_{n=0}^{\infty} \Psi_n y_{t-n}.$$

This formula is interpreted that ξ_t can be decomposed into the previous value of y_t and an innovation. An innovation cannot be explained by past observations and thus is independent from the past.

Here is a simple and clear example. An MA(1) is defined as

$$y_1 = \xi_1 - \theta \xi_0.$$

We repeatedly substitute for lagged values of ξ_t, such as

$$\begin{aligned} y_1 &= \xi_1 - \theta \xi_0 \\ y_2 &= \xi_2 - \theta \xi_1 = \xi_2 - \theta(y_1 - \theta \xi_0) \\ &\cdots \\ y_n &= \xi_n - \theta \xi_{n-1} = \sum_{m=1}^{\infty} (-\theta)^m y_{n-m} + \theta^m \xi_0 + \xi_n. \end{aligned}$$

y_n is expressed as an AR(∞) if $|\theta| \leq 1$. As m increases, $\theta^m \xi_0$ becomes smaller. In other words, an MA(1) is inverted to AR(∞). We can see the invertibility of MA(1) under the restriction of $|\theta| \leq 1$. If we use the property of autocorrelation, we can easily identify the invertibility of the model. As

$$\begin{aligned} E[y_t^2] &= E[(\xi_t - \theta\xi_{t-1})^2] = \sigma^2(1+\theta^2), \\ E[y_t y_{t-1}] &= E[(\xi_t - \theta\xi_{t-1})(\xi_{t-1} - \theta\xi_{t-2})] = -\theta\sigma^2, \end{aligned} \tag{12}$$

we can get the autocorrelation function:

$$\rho(\tau) = \begin{cases} -\theta/(1+\theta^2), & \tau = 1, \\ 0, & \tau \geq 2. \end{cases} \tag{13}$$

If we replace the parameter θ by $1/\theta$, then

$$\rho(1) = \frac{-1/\theta}{1+(1/\theta)^2} = \frac{-\theta}{1+\theta^2}.$$

This implies that when the condition is met, the MA(1) is invertible. If a process is not invertible, one can generate a time series that has an identical autocorrelation structure. But if a process is invertible, such a series has only one autocorrelation structure.

Please note that the condition of invertibility is the same as that of stationarity. MA(2) process is defined as

$$y_t = \xi_t - \theta_1\xi_{t-1} - \theta_2\xi_{t-2},$$

where θ_1 and θ_2 are parameters. As

$$\begin{aligned} E[y_t^2] &= E[(\xi_t - \theta_1\xi_{t-1} - \theta_2\xi_{t-2})^2] = \sigma^2(1+\theta_1^2+\theta_2^2), \\ E[y_t y_{t-1}] &= \sigma^2(-\theta_1 + \theta_1\theta_2), \\ E[y_t y_{t-2}] &= -\theta_2\sigma^2, \end{aligned}$$

we can get the autocorrelation function:

$$\rho(\tau) = \begin{cases} (-\theta_1 + \theta_1\theta_2)/(1+\theta_1^2+\theta_2^2), & \tau = 1, \\ -\theta_2/(1+\theta_1^2+\theta_2^2), & \tau = 2, \\ 0, & \tau \geq 3. \end{cases} \tag{14}$$

The invertibility condition can be shown to be

$$\begin{aligned} \theta_1 + \theta_2 &< 1, \\ -\theta_1 + \theta_2 &< 1, \\ -1 < \theta_2 &< 1. \end{aligned} \tag{15}$$

3.3.3 The exponential weighted moving average

The simplest way to estimate the level of time series is to get the sample mean of the observations. If more weight is given to the most recent sample, then such weight exponentially decays.

$$m_t = (1 - \lambda)m_{t-1} + \lambda y_t,$$

where m represents the current level of the observations. This recursion is an exponentially weighted moving average (EWMA). The optimum λ is obtained by minimizing the sum of squares of the one-step-ahead forecast errors.

Hold (1957) and Winters (1960) extended the model by adding a slope b. It becomes

$$\begin{aligned} m_t &= \lambda_0 y_t + (1 - \lambda_0)(m_{t-1} + b_{t-1}), \\ b_t &= \lambda_1(m_t - m_{t-1}) + (1 - \lambda_1)b_{t-1}. \end{aligned}$$

These equations are known as Holt's recursions. Minimizing the sum of squares of the one-step-forward forecast errors optimizes both λ_0 and λ_1.

3.3.4 The local level model

From equations (2) and (3), we can get the single equation form of the model

$$y_t - y_{t-1} = \Delta y_t = \eta_t + \epsilon_t - \epsilon_{t-1} = \eta_t + \Delta\epsilon_t, \tag{16}$$

where Δ is the first-difference operator. When $\sigma_\epsilon^2 = 0$, it is rewritten as

$$y_t - y_{t-1} = \Delta y_t = \eta_t.$$

The model is said to be the ARIMA(0,1,0) without any restriction.

From equation (16), it is easy to see that

$$\begin{aligned} &E(\Delta y_t) = E(\eta_t) + E(\epsilon_t) - E(\epsilon_{t-1}) = 0, \\ &E[(\Delta y_t)^2] = E[(\eta_t + \Delta\epsilon_t)^2] = \sigma_\eta^2 + 2\sigma_\epsilon^2, \\ &E[(\Delta y_t)(\Delta y_{t-1})] = E[(\eta_t + \Delta\epsilon_t)(\eta_{t-1} + \Delta\epsilon_{t-1})] = -\sigma_\epsilon^2. \end{aligned} \tag{17}$$

Thus we can get the autocorrelation function of the model:

$$\rho(\tau) = \begin{cases} -\sigma_\epsilon^2/(\sigma_\eta^2 + 2\sigma_\epsilon^2), & \tau = 1, \\ 0, & \tau \geq 2. \end{cases} \tag{18}$$

The range of the autocorrelation is given by $-0.5 \leq \rho(1) \leq 0$. It exhibits the cut-off at lag one which is the characteristics of a first-order moving average, MA(1), process.

Taken together equation (18) and (13) yield

$$\theta^2 + (q + 2)\theta + 1 = 0.$$

The solution is given by

$$\theta = \left(\sqrt{q^2 + 4q} - 2 - q\right)/2.$$

It confines the range of $-1 \leq \theta \leq 0$ because of $0 \leq q \leq \infty$. Under the range of θ, the reduced form is invertible. Equation (17) and (12) gives $\theta = -\sigma_\epsilon^2/\sigma^2$. Then the local level model is reduced to the ARIMA(0,1,1) with the restriction of $-1 \leq \theta \leq 0$. The parameter space is half of the one given by the unrestricted reduced form. In the case of $\sigma_\eta^2 = 0$,

$$\theta^2 + 2\theta + 1 = (\theta + 1)^2 = 0,$$

thus $\theta = -1$. The reduced form model is strictly noninvertible.

In relation to the EWMA, optimal forecast of the model is given by an EWMA form. Please see Muth (1960).

3.3.5 The local linear trend model

Taken together equations (9), (10) and (11) yield the single equation form of the model

$$\Delta^2 y_t = \Delta\eta_t + \lambda_{t-1} + \Delta^2\epsilon_t. \tag{19}$$

When $\sigma_\lambda^2 = 0$, it is rewritten as

$$\Delta y_t = \eta_t + \lambda + \Delta\epsilon_t.$$

The model is said to be ARIMA(0,1,1) with the same restriction for the local level model.

From equation (19), it can be seen that

$$\begin{aligned}
&E[(\Delta^2 y_t)^2] = E[(\Delta\eta_t + \lambda_{t-1} + \Delta^2\epsilon_t)^2] = 2\sigma_\eta^2 + \sigma_\lambda^2 + 6\sigma_\epsilon^2, \\
&E[(\Delta^2 y_t)(\Delta^2 y_{t-1})] = -\sigma_\eta^2 - 4\sigma_\epsilon^2, \\
&E[(\Delta^2 y_t)(\Delta^2 y_{t-2})] = \sigma_\epsilon^2.
\end{aligned}$$

Thus we can get the autocorrelation function of the local linear trend model:

$$\rho(\tau) = \begin{cases} (-\sigma_\eta^2 - 4\sigma_\epsilon^2)/(2\sigma_\eta^2 + \sigma_\lambda^2 + 6\sigma_\epsilon^2), & \tau = 1, \\ (\sigma_\epsilon^2)/(2\sigma_\eta^2 + \sigma_\lambda^2 + 6\sigma_\epsilon^2), & \tau = 2, \\ 0, & \tau \geq 3. \end{cases} \tag{20}$$

It exhibits the cut-off at lag two that is characteristic of a second-order moving average, MA(2), process. These variances confine the range of the autocorrelations:

$$\begin{aligned}
-0.667 \leq \rho(1) \leq 0, \\
0 \leq \rho(2) \leq 0.167.
\end{aligned}$$

When $\sigma_\eta^2 = 0$ and $\sigma_\lambda^2 = 0$, $\rho(1) = -0.667$ and $\rho(2) = 0.167$, and $\sigma_\eta^2 = 0$ and $\sigma_\epsilon^2 = 0$, then $\rho(1) = 0$. When $\sigma_\epsilon^2 = 0$, then $\rho(2) = 0$.

The local linear trend model is reduced to the ARIMA(0,2,2). However, from equation (15) and the given ranges of autocorrelations, the restrictions applied to this model are much narrower than those applied to the local level model. Goddphin and Stone (1980) provide the detail analysis.

Optimal forecast in the model is exactly same as that of Holt's recursions.

3.3.6 The higher order polynomial trends

The unobserved components form of the model is given by

$$\Delta^3 y_t = \Delta^2 \eta_t + \Delta(\lambda_{t-1} + 0.5\psi_{t-2}) + \Delta^2 \epsilon_t.$$

The model is reduced to the ARIMA(0,3,3). Please see the discussion in Jacobs and Jones (1980).

3.3.7 Reduced form parameters

We have looked at several state space, EWMA and ARIMA models. The state space model can be reduced to the ARIMA form with certain restrictions. The EWMA models also have similar properties to the corresponding state space models. We can find some discrepancies between the state space form and the corresponding ARIMA model in terms of the admissible region of reduced form parameters. State space form can be reduced to the ARIMA class models but have certain restrictions. These restrictions are not a disadvantage of state space form. Rather, they can be a powerful tool to generate and control data consistent with a given sample.

4 The time-varying models

Analysis of the local level and the local linear trend models that have several disturbance terms shows some flexibility in the time invariant state space form. We reduce them to the ARIMA models with a single disturbance term and stationarity. There are some limitations in the ARIMA framework and some potential in the state space form for handling complex variation in financial time series. There is more flexible way to develop the state space model for a wide variety of series complexity: the system matrices can change their values in some predetermined manner when certain conditions are met. Such models are said to be time-varying.

In structural time series models, each component evolves stochastically over time under the assumption of fixed hyperparameters. Each component changes the values of its parameters consistently with given observations. These models can be reduced to stationarity by some transformations. However maintaining stationarity is not always an essential part of financial time series modeling. If we realize that the hyperparameters are not optimal for the given data, we have

to calibrate and change them. The fixed hyperparameters are time-varying. The model with time variation may lose its linearity and so its stationarity and be classified as a nonlinear model.

We analyze three classes of time-varying models: the conditionally Gaussian models, the regime switching models, and the explanatory variables with time-varying coefficients. These models are nonlinear, however their hyperparameters may shift in order to keep the process stationary. As a result of such characteristics, they are distinguished from the functionally nonlinear models.

4.1 Conditionally Gaussian models

The type of distribution of disturbances may characterize the property of the model. Many of disturbances are assumed to follow a Gaussian distribution. Conditionally Gaussian models have Gaussian disturbances, and the system matrices evolve as a function of given observations at time t. The values of hyperparameters change as the system matrices evolve with the passing of time. Such models are specified as the following set of equations:

$$\begin{aligned} \mathbf{y_t} &= \mathbf{Z_t}(\mathbf{Y_{t-1}})\alpha_\mathbf{t} + \mathbf{d_t}(\mathbf{Y_{t-1}}) + \epsilon_\mathbf{t}, \\ \alpha_\mathbf{t} &= \mathbf{T_t}(\mathbf{Y_{t-1}})\alpha_\mathbf{t-1} + \mathbf{c_t}(\mathbf{Y_{t-1}}) + \mathbf{R_t}(\mathbf{Y_{t-1}})\eta_\mathbf{t}, \end{aligned}$$

where

$$\begin{aligned} &\epsilon_\mathbf{t} \mid \mathbf{Y_{t-1}} \sim \mathbf{N}[\mathbf{0}, \mathbf{H_t}(\mathbf{Y_{t-1}})], \\ &\eta_\mathbf{t} \mid \mathbf{Y_{t-1}} \sim \mathbf{N}[\mathbf{0}, \mathbf{Q_t}(\mathbf{Y_{t-1}})] \quad \textbf{, and} \\ &\alpha_\mathbf{0} \sim \mathbf{N}(\mathbf{a_0}, \mathbf{P_0}). \end{aligned}$$

The system matrices may change as a function of given observations. However the hyperparameters are regarded as being fixed at that very moment that the matrices evolve. Therefore, we can use the Kalman filter to derive the general linear Gaussian state space models. Lipster and Shirayav (1978) have undertaken the extensive research in this field .

4.1.1 An AR(1) model with parameter following AR(1) process

Weiss (1985) discussed the stability of an AR(1) model with a time-varying coefficient.

$$\begin{aligned} y_t &= (\psi_0\psi_t)y_{t-1} + \epsilon, \quad \epsilon \sim N(0,1), \\ \psi_t &= M\psi_{t-1} + \eta_t, \quad \eta \sim N(0,Q), \end{aligned}$$

where $E(\epsilon_t\eta_s) = 0$ for all t and s and $\mathrm{Var}(\epsilon_t) = 1$ and $\mathrm{Var}(\eta_t) = Q$. The interaction between the two equations is dynamic and very complicated. Setting up $M = 0$ and $Q = 0$ reduces the above state space form to the usual AR(1) process. The condition of stability is $|\psi_0| < 1$. Weiss makes the interesting observation that under certain conditions the process is stable even though the value of ψ may take the explosive region $|\psi| \geq 1$.

4.2 Structural Break

It is assumed that a financial time series is reasonably stable, and that when its characteristics change, the process goes slowly. However sometimes the characteristics shift suddenly from one condition to another. These phenomena are known as structural breaks.

Stationarity is an essential element of structural time series modeling. Each component evolves stochastically over time and changes the values of its parameters consistent with given observations. The series is modeled as nonstationary, but can be reduced to stationarity, therefore these hyperparameters are fixed in each regime. If the property of stationarity does not fit the stylized fact, we have to drop it. This is the basic notion of modeling structural break.

4.2.1 Detecting structural break

Structural changes have been researched for more than fifty years since Chow (1960) developed a test to detect structural shift. Many advances have been made to cover practical applications in the context of

- known/unknown change points: Zivot and Andrews (1992), Hansen (1992)
- single/multiple structural changes: Bai (1999), Bai and Perron (1998), Hansen (2000)
- endogenous/endogenous model: Perron (1989), Zivot and Andrews (1992)
- threshold/stochastic process: Perron (1989)
- permanent/temporal breaks: Rappoport and Reichlin (1989), Perron (1989)

The most work has been devoted to detecting structural breaks.

4.2.2 Switching regime

The observations in a time series may not be generated by the same mechanisms over time, that is, the mechanisms may change at different points in time. In that case, the series is subject to switching regimes.

Construct the two-regime switching model that is modeled by two local level models. Two regimes Ψ_1 and Ψ_2 represent the normal financial markets and the markets under systemic risk, respectively. That is $\Psi_1 = [\sigma^2_{\epsilon_1}, \sigma^2_{\eta_1}]$ and $\Psi_2 = [\sigma^2_{\epsilon_2}, \sigma^2_{\eta_2}]$. These hyperparameters should be estimated separately based on the past observations of normal markets and the markets exposing the systemic risk, if we know when the change in regime has taken place. If a regime change is generated endogenously, the model becomes nonlinear. If the following conditions are met, the model might be conditionally Gaussian:

- a finite number of regimes,
- each regime has a different set of hyperparameters,

- switching regimes are determined by a function of the past observations or threshold model.

Optimization of each hyperparameter is undertaken separately, therefore each local level model is obviously linear.

4.3 Explanatory variable

Equation (7) includes the explanatory variable d_t. However if the regression coefficient of the explanatory variable is constant, it becomes just normal regression problem. d_t is replaced by $\delta_t x_t$, where δ_t evolves stochastically, and x_t is an explanatory variable. The estimation of structural time series models with explanatory variables follows the process applied to the general linear Gaussian models.

4.3.1 The time-varying CAPM

The time-varying CAPM is formed as a conditionally Gaussian model. It is given by

$$\begin{aligned} E(R_t) - R_f &= \alpha_t + \beta_t[E(R_{M,t}) - R_f] + \epsilon_t, \\ \alpha_{t+1} &= \alpha_t + \eta_t, \\ \beta_{t+1} &= \beta_t + \lambda_t, \end{aligned}$$

where $E(R_t)$ is the expected return of an asset, R_f is the risk-free rate of interest, $E(R_{M,t})$ is the expected return on the market portfolio, β_t is the sensitivity of the expected excess asset return to the expected excess market returns. ϵ_t, η_t and λ_t are disturbance terms, uncorrelated with each other and uncorrelated with the initial state over time. α and β are assumed to evolve stochastically. In the original CAPM, $\alpha = 0$ and β is constant. Please see Tsay (2005.p 510).

4.4 Stochastic volatility model

The most important application of state space modeling to financial time series is a stochastic volatility model. This approach was proposed by Hull and White (1987), Chesney and Scott (1989), and Melion and Turnbull (1990). Their work discussed time-varying volatility in an option pricing model. The Black-Scholes (BS) model assumes that volatility is constant over time to maturity. However, this work derived more dynamic BS type option pricing formulas. A simple example is

$$\begin{aligned} y_t &= \epsilon_t \exp(h_t/2), \epsilon_t \sim NID(0,1), \\ h_t &= \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \eta_t \sim NID(0, \sigma_h^2). \end{aligned}$$

One of the disadvantages of this model is that it requires computer-intensive methods such as MCMC and importance sampling. However, recent development of computational power enables these methods to be less time consuming.

Lam and Li (1998) propose a regime-switching stochastic volatility model in which the model has three states. They apply the model to S&P weekly returns.

Harvey, Ruiz and Shephard (1991) examine four daily exchange rates and suggest that all the rates followed a random walk.

5 Prediction

Prediction can be defined as making a statement on what or how an object will be in the future. Sometime prediction is a major objective for a time series modeling. In statistics, the prediction of y_t for l steps ahead is defined to get the optimal predictor of y_{t+l}, that is the expected value of y_{t+l} conditional on given information at time t. We denote the optimal value of y_{t+l} on given information Y_t at time t as $\tilde{y}_{t+l|t}$. The estimation error is decomposed into

$$y_{t+l} - \tilde{y}_{t+l|t} = [y_{t+1} - E(y_{t+l} \mid Y_t)] + [E(y_{t+l} \mid Y_t) - \tilde{y}_{t+l|t}].$$

Squaring and taking conditional expectations for the whole equation yields

$$\begin{aligned} \text{MSE}(\tilde{y}_{t+l|t}) &= E(y_{t+l} - \tilde{y}_{t+l|t})^2 \\ &= E[y_{t+1} - E(y_{t+l} \mid Y_t)]^2 + E[\tilde{y}_{t+l|t} - E(y_{t+l} \mid Y_t)]^2 \quad (21) \\ &= \text{var}(y_{t+l} \mid Y_t) + \text{bias}(\tilde{y}_{t+l|t})^2. \end{aligned}$$

In case of an AR(1) process, $y_t = \phi y_{t-1} + \xi_t$, it is rewritten as

$$y_t = \sum_{m=o}^{m} \phi^m \xi_{t-m}.$$

- $\phi = 1$, it becomes the random walk and the forecast function becomes horizontal.
- $|\phi| < 1$, the forecasts decay exponentially toward zero. The process is stationary and invertible.
- $|\phi| > 1$, the forecasts are explosive and noninvertible.

The optimal predictors of y_{t+l} at time t are the expected value of y_{t+l} estimated at time t with a given observation up to time t:

$$\tilde{y}_{t+l|t} = E(y_{t+l} \mid Y_t).$$

This predictor is optimal when it has a minimum mean square error (MMSE), given by equation (21).

The optimal predictors of the AR(1) process are expressed as

$$\tilde{y}_{t+l|t} = \phi^l y_t,$$

if the disturbances are independent, rather than uncorrelated, otherwise it is in the class of a linear predictor. This expression is known as the forecast function.

5.1 One step and multi-step prediction

In the general linear Gaussian state space model, the time-invariant Kalman filter yields a prediction one step ahead, a_{t+1}. It is expressed by equation (8). The optimal prediction of y_{t+1} at time t can be

$$E(\mathbf{y_{t+1}} \mid \mathbf{Y_t}) = \mathbf{Z a_{t+1}} + \mathbf{d_{t+1}}.$$

Now consider the multi-step prediction, that is, predicting future observations more than one step ahead. Substituting repeatedly in the prediction equations (8), step by step, up to $t + l$ yields

$$\mathbf{a_{t+l}} = \mathbf{E}(\alpha_{\mathbf{t+l}} \mid \mathbf{Y_t}) = \left[\prod_{\mathbf{j=1}}^{\mathbf{l}} \mathbf{T_{t+j}}\right] \mathbf{a_t} + \sum_{\mathbf{j=1}}^{\mathbf{l-1}} \left[\prod_{\mathbf{i=j+1}}^{\mathbf{l}} \mathbf{T_{t+i}}\right] \mathbf{c_{t+j}} + \mathbf{c_{t+l}}.$$

Its covariance matrix is

$$\mathbf{P_{t+l}} = \mathbf{E}[\mathrm{Var}(\alpha_{\mathbf{t+l}} \mid \mathbf{Y_t}) = \mathbf{T^l P_t T^{l\prime}} + \sum_{\mathbf{j=0}}^{\mathbf{l-1}} \mathbf{T^j R Q R' T^{j\prime}}.$$

For a time-invariant model of the form, the multi-step predictor is

$$\mathbf{y_{t+l}} = \mathbf{Z a_{t+l}} = \mathbf{Z T^l a_t}.$$

5.2 Prediction under time-variation

In time-variant models, MMSEs of future observations are not always guaranteed. The characteristics of the model and the range of parameters affect the properties of steadiness, stationarity, and invertibility.

5.2.1 Conditionally Gaussian

Since the models are nonlinear, the MMSEs of future observations are not always guaranteed. They may or may not be linear depending on the basic properties of the models and the values of some parameters.

5.2.2 Regime switching

If the observations are normal in each regime, it may possible that the optimal estimators are MMSEs.

5.2.3 Explanatory variable with time-varying coefficients

The linearity in the MMSEs of future observations may or may not depend on the basic properties of the transition equations and the values of some parameters in the model.

5.3 Post sample prediction test

Due to complicated properties of financial time series data, a single measure of forecasting performance analysis may not be sufficient. Four types of statistical methods are available to evaluate forecast performance:

- The absolute predictive accuracy
  - Mean Square Error : MSE $= E[(y_t - a_t)^2]$
  - Mean Error : ME $= \frac{1}{n}\sum(y_t - a_t)$
  - Mean Percentage Error : MPE $= \frac{1}{n}\sum[(y_t - a_t)/y_t]$
  - Mean Absolute Error : MAE $= \frac{1}{n}\sum(|y_t - a_t|)$
  - Mean Absolute Percentage Error : MAPE $= \frac{1}{n}\sum[|(y_t - a_t)/y_t|]$
  - The Theil's U Statistic
- The directional predictive ability
  - The Henriksson-Merton test
  - The contingency table analysis
- The distributional measure
  - The Kolmogorov-Smirnov test
- The entropic measure
  - The Kullback-Leibler distance

where a_t is the optimal predictor of y_t.

The weakness of the absolute predictive measures is that the results are affected by outliers or structural breaks. This weakness can be addressed by implementing rolling duration or some optimization of a time horizon. On the other hand, the directional measures do not have that weakness. Outlier and structural break may not have a strong cumulative impact on the results. However they only measure the direction.

5.3.1 The mean squared error

The mean squared error is a measure of how close predictions are to actual outcomes. It is defined as the expected value of the squared errors; that is,

$$\text{MSE} = E[(y_t - a_t)^2],$$

where a_t is the expected value of y_t at time $t-1$ at given information Y_{t-1}. On the other hand, the variance of y_t is defined as

$$\text{var}(y_t) = E[y_t - E(y_t)]^2.$$

This is a parameter measuring how far variables lie from the expected value. It describes either the actual value or theoretical value. If these variables are not fully observable, the sample data is used to estimate its variance. It is called the sample variance. The relationship between the mean squared error and the variance is described as

$$\mathrm{MSE} = \mathrm{var}(y_t) + \mathrm{bias}(a_t)^2.$$

If there is no bias, then $\mathrm{var}(y_t) = \mathrm{MSE}(a_t)$.

Please note that the MSE is an expectation, therefore it is a scalar and not a random variable. As a result of squaring of each term, the MSE weights large errors more heavily than small ones; this property becomes a disadvantage in many applications. In this case, the mean absolute error is one of alternatives.

5.3.2 The Theil's U statistic

Theil's U statistic is widely used to evaluate forecasts. Two types of U statistic are available for the measures of forecast performance. One shows how close actual and forecast series are; and the other compares the forecast accuracy of the model with naive forecasts.

The U1 statistic is obtained by

$$\mathrm{U1} = \frac{\sqrt{\sum_{t=1}^{n}(\bar{y}_t - y_t)^2}}{\sqrt{\sum_{t=1}^{n}\bar{y}_t^2} + \sqrt{\sum_{t=1}^{n}y_t^2}},$$

where $\bar{y}_t$ is the mean of y_t. U1 is a range between 0 and 1. The closer the value is to zero, the greater the forecasting accuracy is.

The U2 statistic is

$$\mathrm{U2} = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{y_{t+1}-\bar{y}_{t+1}}{\bar{y}_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{\bar{y}_{t+1}-\bar{y}_t}{\bar{y}_t}\right)^2}}.$$

U2 = 1 indicates no difference between the naive forecast and the forecast from the model.

5.3.3 The Henriksson-Merton test(HM) and the contingency table χ^2 analysis

Henriksson and Merton (1981) proposed a measure of directional predictive ability in which the direction of a forecast generated by a given model is compared with the direction of real price movement. The HM test introduces the 2×2 contingency Table 1. The HM test statistics is given by

$$HM = \frac{n_{11} - \frac{n_{10}n_{01}}{n}}{\sqrt{\frac{n_{10}n_{01}n_{20}n_{02}}{n^2(n-1)}}} \sim N(0,1).$$

		actual		
		up	down	
predicted	up	n_{11}	n_{12}	n_{10}
	down	n_{21}	n_{22}	n_{20}
		n_{01}	n_{02}	n

Table 1: Contingency table

We can apply the Chi-square test to the 2×2 contingency table (Please see Table1). Suppose that certain events are observed and summarized in ith bins and their expected number is drawn from some known distribution. The number of actual events in the ith bin and its expected number are assumed to be n_i and m_i, respectively. Then the Chi-square statistics are given by

$$\chi^2 = \sum_i^I \frac{(n_i - m_i)^2}{m_i},$$

where I is the total number of bins.

We can extend this analysis to the 2×2 contingency table. The Chi-square statistic is

$$\chi^2 = \sum_{1=1}^{2}\sum_{j=1}^{2} \frac{\left(n_{ij} - \frac{n_{i0}n_{0j}}{n}\right)^2}{\frac{n_{i0}n_{0j}}{n}} \sim \chi^2(1),$$

where $\chi^2(1)$ is a χ^2 distribution with one degree of freedom. The null hypothesis is that the volatility of actual data is independent of the volatility of predictions.

- $H_0 : p_{up}^{01} = p_{up}^0 p_{up}^1$
- H_1: $p_{up}^{01} \leq p_{up}^0 p_{up}^1$,

where p_{up}^{01} is the probability of actual and predicted value "up." p_{up}^0 and p_{up}^1 represent the probability of actual value up and the probability of predicted value up, respectively. As you can see, the Chi-square statistic is larger as the n_{11} and n_{22} or the n_{12} and n_{21} dominate the events. The Chi-square does not show which has happened. Therefore we need another measure to identify the predictability such as the confusion rate CR:

$$CR = \frac{n_{12} + n_{21}}{n}.$$

When $CR > 0$ then

- $H_0 : \chi^2 < \chi_0^2$
- $H_1 : \chi^2 \geq \chi_0^2$.

The statistic is larger than the critical value, and the null is rejected.

5.3.4 The Kolmogorov-Sminov test

When a set of realizations is continuous and has a one dimensional distribution, the Kolmogorov-Sminov (KS) test is applicable as a nonparametric test to compare a sample with the normal distribution. The overall difference between the cumulative distribution function(cdf) of the given sample and the cdf of the normal distribution can be measured by the absolute value of the area between them. The KS test uses the maximum value of the absolute difference between two cumulative distribution functions. The KS statistic is defined as

$$D = \max |f(x) - p(x)|,$$

where $f(x)$ and p(x) are the cdf of the normal distribution and the cdf of a given empirical distribution, respectively. The p-value of an observed value of D is obtained from

$$Q_{\mathrm{KS}}([\sqrt{T} + 0.12 + 0.11/\sqrt{T}]D),$$

where T is the sample size. It follows the KS distribution given by

$$\begin{aligned} P_{\mathrm{KS}}(z) &= 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2z^2), \\ &= \frac{\sqrt{2\pi}}{z}\sum_{j=1}^{\infty}\exp -(\frac{(2j-1)^2\pi^2}{8z^2}), \end{aligned}$$

and

$$Q_{\mathrm{KS}}(z) = 1 - P_{\mathrm{KS}}(z),$$

for all $z > 0$.

- H_0 : the disturbances are normally distributed
- H_1 : the disturbances are not normally distributed.

These are restated as:

- H_0 : $z \leq z_0$
- H_1 : $z > z_0$.

This is a two-sided test.

Alternative tests are the Kuiper's test, the Cramer-von Mises test, and the Anderson-Darling test.

5.3.5 The Kullback-Leibler distance

We can define entropy of a distribution p,

$$H(p) = -\sum_{i=0}^{I-1} p_i \ln p_i,$$

where p denotes a distribution with outcome i and its associated probability p_i. The entropy is zero when one of the p_i is unity and all other outcomes are zero. It takes on the maximum value when all p_i are the same.

Kullback Leibler distance is defined as

$$D(p;q) = \sum p_i \ln \frac{p_i}{q_i},$$

where p is a distribution and q is some other distribution. It is also called the relative entropy. Using the fact of $\ln a \leq a - 1$, we can prove its nonnegativity,

$$-D(p;q) = \sum p_i \ln \frac{q_i}{p_i} \leq \sum p_i \left(\frac{q_i}{p_i} - 1 \right) = 1 - 1 = 0.$$

If two distributions are identical, then the Kullback Leibler distance between two distributions is zero.

Let us consider that the random variable X follows the binomial distribution with parameters n and q in a coin toss: n is the total number of trials; q is the probability of success in n independent trials. The probability of getting k successes in n trials is given by

$$\Pr(X = k) =_n C_k q^k q^{n-k},$$

for $k = 0, 1, 2, \ldots, n$, where

$$_nC_k = \frac{n!}{k!(n-k)!}.$$

This is the binomial coefficient when n chooses k. We can use Stiring's formula for large factorials: $M! = \sqrt{2\pi M}(m/e)^M$ and rewrite the above formula:

$$\begin{aligned} \Pr(X = k) &= \exp n[-p\ln(p) - (1-p)\ln(1-p) + p\ln(q) + (1-p)\ln(1-q)] \\ &= e^{-nD(p;q)}, \end{aligned}$$

where $p = k/n$. Please refer to more rigorous discussion in Aoki (1996).

6 Conclusion

We have looked at four types of financial time series modeling: the EWMA, the ARIMA, the time-invariant state space model and the time-variant state space model. The time-variant models are reduced to the ARIMA class models with certain restrictions. These restrictions originated from the invertibility of the ARMA model. They are not disadvantages of the state space model. This shows that the state space model has a more flexible structure, created by the number of disturbance terms. In general, the state space models have more than one disturbance term, while the ARIMA model has only one disturbance.

In addition, the state space models may be a more flexible way to handle the wide variety of complex financial time series. For example, stochastic volatility

models have the potential to replicate a given observation. The regime switching model applied to the stochastic volatility model may generate data more consistent with a given sample in the long-run. When we use these models in applied work, we have to carefully examine the parameter space in which they are in steady state. We have to know that the optimal predictor is just the expected value of y_{t+l} conditional on the given information at time t, and that it only provides the conditional distribution of future observations with the MMSEs.

Measures of prediction performance may or may not be useful depending on how one uses them, and in what kind of situations. We cannot have deterministic solutions.

Prediction is an essential part of financial trading and risk management. However, it is rarely an end in itself and please note optimal predictors are not accurately "fortune telling".

References

[1] Aoki, M.(1996) New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria, and Externalities as Field Effects. Cambridge University Press

[2] Bai,J., (1999) Likelihood ratio tests for multiple structural change, Journal of Econometrics, 91 (299-323)

[3] Bai,J.,and Perron, P.,(1998) Estimating and testing linear models with multiple structural changes. Econometrica, 66(47-78)

[4] Box,J. and G.Jenkins (1970) Time series analysis: forecasting and control, San Francisco: Holden-Day

[5] Chow,G.C. (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions, Econometrica, 28 (591- 605)

[6] Durbin, J. and S.J.,Koopman(2001) Time Series Analysis by State Space Methods, Oxford University Press.

[7] Fletcher, R. (1987) Practical Methods of Optimisation 2nd edition. New York: John Wiley.

[8] Godolphin, E., and M.Stone (1980) On the structural representation for polynomial predictor models. Journal of the Royal Statistical Society, Series B 42(35-45)

[9] Hansen, B.E.(1992) Tests for parameter instability in regressions with I(1) processes. Journal of Business and Economic Statistics, 10 (32-335)

[10] Hansen, B.E.(1997) Testing for structural change in conditional models, Journal of Econometrics, (93-115)

[11] Harvey, A.C.(1989) Forecasting, structural time series models and the Kalman filter, Cambridge University Press

[12] Harvey, A.C. Ruiz,E. and N.G.,Shephard (1994) Multivariate stochastic variance models, Review of Economic Studies, 61, Issue 2(247-264)

[13] Henriksson, R.and R.Merton (1981) On market timing and investment performance .ii. statistical procedures for evaluating forecasting skills,The Journal of Business 54 (513-533)

[14] Holt, C.C. (1957) Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

[15] Jacobs, R.L., and R.A.,Jones (1980) Price expectations in the United States, 1947-75. The American Economic Review, 70(267-77)

[16] Kalman, R.E.(1960) A new approach to linear filtering and prediction problems, J.Basic Engineering, Transactions ASMA, Series D,82(35-45)

[17] Kolmogorov A.N.(1942) Stationary sequences in Hilbert space, (in Russian) Bull. Moscow Univ. , 2 No.6(1-40)

[18] Koopman,S.J.(1993) Disturbance smoother for state space methods, Biometrika, 80(117-26)

[19] Lam, So,M.K.P., and W.K.,Li (1998) A stochastic volatility model with Markov switching, Journal of Business and Economics Statistics, 16(244-53)

[20] Liptser, R.S. and A.N.,Shiryayev (1978) Statistics of Random Processes II: Applications, Springer, New York

[21] Ljung, G.M. and G.E.P.,Box (1978) On a Measure of lack of Fit in Time Series Models, Biometrika, 66(66-72)

[22] Nerlove,M.,D.M. Grether and J.L.Varvalho(1979) Analysis of Economic Time Series. New York: Academic Press

[23] Perron. P., (1989) The great crash the oil price shock and the unit root hypothesis. Econometrica, 57(1361-1401)

[24] Rappoport,P., and L.,Reichlin (1989) Segmented trends and nonstationary time series. The Economic Journal, 99 (Conference Volume)(168-177)

[25] Gonzalez-Rivera G.(1997) The Pricing of Time-Varying Beta, Empirical Economics, 22(345-363)

[26] Shumway, R.H. and D.S.,Stoffer (1982) An approach to time series smoothing and forecasting using the EM algorithm, J. Time Series Analysis, 3(253-64)

[27] Tsay,R. S. (2005) Analysis of Financial Time Series, John Wiley & Sons, Inc.

[28] Watson, M. W. and R.F.,Engle (1983) Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression, Journal of Econometrics, 23(385-400)

[29] Weiss, A.A. (1985) The stability of the AR(1) process with an AR(1) coefficient. Journal of Time Series Analysis, 6:(181-6)

[30] Whittle, P. (1983) Prediction and regulation by linear least-square methods, second edition, revised. University of Minnesota Press.

[31] Wiener, Norbert (1949) Extraporation, Interporation, and Smoothing of Stationary Time Series, New York Wiley

[32] Winters, P.R. (1960) Forecasting sales by exponentially weighted moving averages. Management Science, 6(324-42)

[33] Zivot,E. and D.W.K., Andrews(1992) Further evidence on the great crash, the oil-price shock, and the unit root hypothesis, Journal of Business and Economics Statistics, 10 (251-270)