EN.550.782 Final Project: Dynamic Linear Models Analysis of S&P 500

Shibai Zhang

December 14, 2016

1 Introduction

S&P500 index[1] is the Standard & Poor's 500, often abbreviated as the S&P 500, or just "the S&P", is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ .Analysis of S&P500 has an important meaning because it is seen as a leading indicator of U.S. equities and a reflection of the performance of the large cap universe.

When it comes to the models for time series analysis, the class of autoregressive moving average (ARMA) models and the ARCH(Autoregressive conditional heteroscedasticity) models are most widely used. But in this report, we will apply some models that have not been used in the analysis of S&P 500 index. We will use Dynamic Linear Models, a special case of state-space models to analysis the time series data, so as to check if Dynamic Linear Models works on forcasting S&P 500 index.

2 Basic Theories

2.1 State-space Model

First, we will introduce the structure of a state-space model. Formally, a state-space model consists of an R^p-valued time series (θ_t : t = 0,1 ...) and an R^m-valued time series (Y_t : t = 1,2 ...), satisfying the following assumptions:

(A.1) (θ_t) is a Markov chain, that is θ_t depends on the past values (θ_0, θ_1, ...,θ_{t-1}) only through θ_{t-1}.

(A.2) Conditionally on (θt, t = 0,1 ...), the Y_t's are independent and Y_t depends on θ_t only, which means that (Y_1, ..., Y_n)|θ_1, ...,θ_n have joint conditional

[1] wikipedia, https://en.wikipedia.org/wiki/S&P_500_Index

density $\prod_{t=1}^{n} f(y_t|\theta_t)^1$.

State-space models have been put into use of the time series analysis for years. When we analyze a time series data with State-space models, the data is considered as the output of a dynamic system perturbed by random disturbances. And with state-space models, we could interpret a time series as the combination of several components, such as trend, seasonal or regressive components. At the same time, they have an elegant and powerful probabilistic structure, offering a flexible framework for a very wide range of applications. For example, the popular ARMA model can also be represented into a state-space model as a special case. We will interpret it specifically in the latter part. In state-space models, with the given information, the problems of estimation and forecasting are solved by recursively computing the conditional distribution. State space models can be used to model univariate or multivariate time series, also in presence of non-stationarity, structural changes, irregular patterns. Here in the case of S&P500 index, it is a univariate time series data.

2.2 Dynamic Linear Models

Dynamic linear models, which are given by Gaussian linear state-space models, is a special case of state-space models.

A dynamic linear model (DLM) can be written regularly as the following form,

$$\theta_0 \sim MN(m_0, C_0) \tag{1}$$

$$Y_t = F_t\theta_t + v_t, v_t \sim i.i.d.N(0, V_t) \tag{2}$$

$$\theta_t = G_t\theta_{t-1} + w_t, w_t \sim i.i.d.MN(0, W_t) \tag{3}$$

where θ_0 is independent of (V_t) and (W_t), G_t and F_t are known matrices, (v_t) and (w_t) are two independent white noise sequences with mean zero and known variance matrices (V_t) and (W_t).

Usually, equation (2) is called the observation equation and (3) is called the state equation or system equation.

We are familiar with some of the special cases in the dynamic linear models, such as the Random walk plus noise model, Discretized Ornstein-Uhlenbeck process, Linear growth model and so on. We will apply the Random walk plus noise model in the latter part.

3 Data

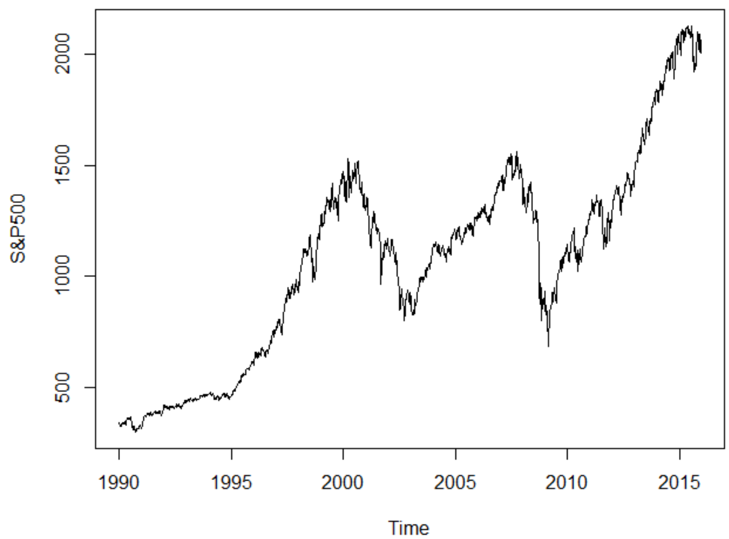
We take the original data from the Yahoo finance[1]. We take the weekly closing data, and the time period is from 1990-1-3 to 2015-12-25, 1352 index

[1] Giovanni Petris, Sonia Petrone, Patrizia Campagnoli, *Dynamic LinearModels with R* ,Springer Dordrecht Heidelberg London New York

[1] https://finance.yahoo.com/quote/%5EGSPC?p=GSPC

values in total.

The plot below is the weekly closing values for the S&P 500 stock market index from 1990 to 2015.



We can see that during the past 25 years, the S&P 500 index increases generally with two big drops.

4 Models Selection

We now apply dynamic linear models in two specific models, Random walk plus noise and ARMA model. We first start with the simplier one.

4.1 Random walk plus noise

We will try first with a simple dynamic linear model. The random walk plus noise model. In this situation,

$$Y_t = \mu_t + v_t, v_t \sim i.i.d.N(0, V) \tag{4}$$

$$\mu_t = \mu_{t-1} + w_t, w_t \sim i.i.d.MN(0, W) \tag{5}$$

Obviously, this is a DLM regular form with $\theta_t = \mu_t$, and $F_t = G_t = 1$

4.2 ARMA model

And the we comes to a more specific model of Dynamic linear models, ARMA models. ARMA models are widely used in time series analysis and here we try it with the DLMs. And ARMA model can be presented in a DLM form.

Here is the definition of regular ARMA models,

$$Y_t = \mu + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu) + \sum_{j=1}^{q} \psi_j \epsilon_{t-j} + \epsilon_t \tag{6}$$

where p and q are nonnegative integers, and (ϵ_t) is Gaussian white noise with variance σ_ϵ^2 and the parameters $\phi_1, ..., \phi_p$ satisfy a stationarity condition.

Here DLM representation of ARMA,

For simplicity, we assume that μ is zero, then present function (6) as following,

$$Y_t = \sum_{j=1}^{r} \phi_j Y_{t-j} + \sum_{j=1}^{r-1} \psi_j \epsilon_{t-j} + \epsilon_t \tag{7}$$

where $r = \max\{p, q+1\}$, $\phi_j = 0$ for $j > p$ and $\psi_j = 0$ for $j > q$.

With matrices,

$$F = \begin{pmatrix} 1 & 0 & ... & 0 \end{pmatrix}$$

$$G = \begin{pmatrix} \phi_1 & 1 & 0 & ... & 0 \\ \phi_2 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ \phi_{r-1} & 0 & ... & 0 & 1 \\ \phi_r & 0 & ... & 0 & 0 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & \psi_1 & ... & \psi_{r-2}\psi_{r-1} \end{pmatrix}^T$$

If one introduces an r-dimensional state vector $\theta_t = (\theta_{1,t},...,\theta_{r,t})^T$, then the ARMA model can be represented as the following dynamic linear model,

$$Y_t = F\theta_t \tag{8}$$

$$\theta_{t+1} = G\theta_t + R\epsilon_t \tag{9}$$

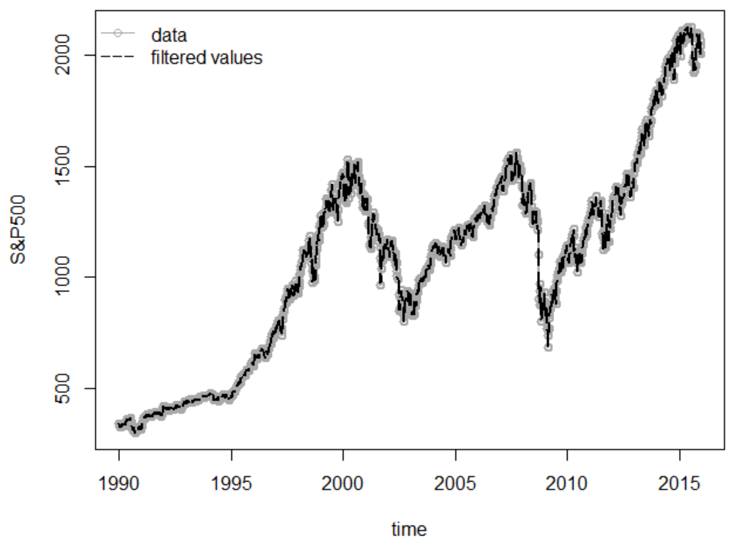
which is a dynamic linear model with $V = 0$ and $W = RR^T\sigma^2$, where σ^2 is the error sequence (ϵ_t).

5 Results

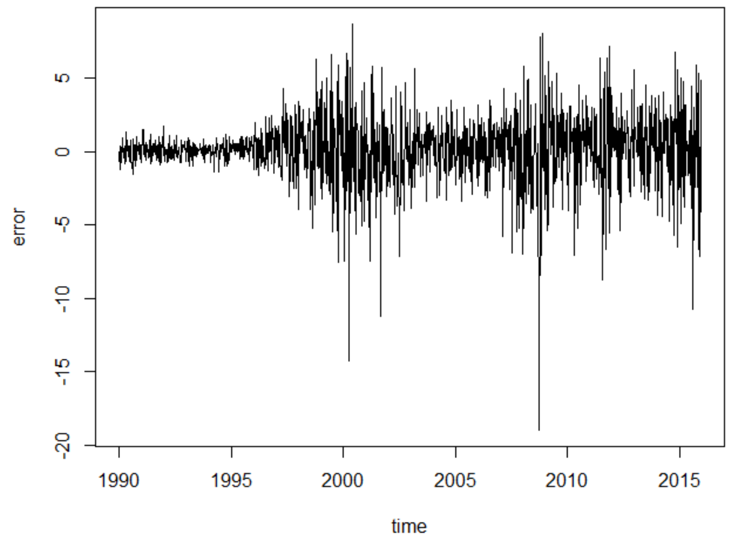
5.1 Random walk plus noise

We use dlm package in R. And in this random walk plus noise model we get the value of V = 60.51726 and W = 557.41563 through the maximum likelihood estimates. And we set μ_0 the mean of the total 1352 index values.

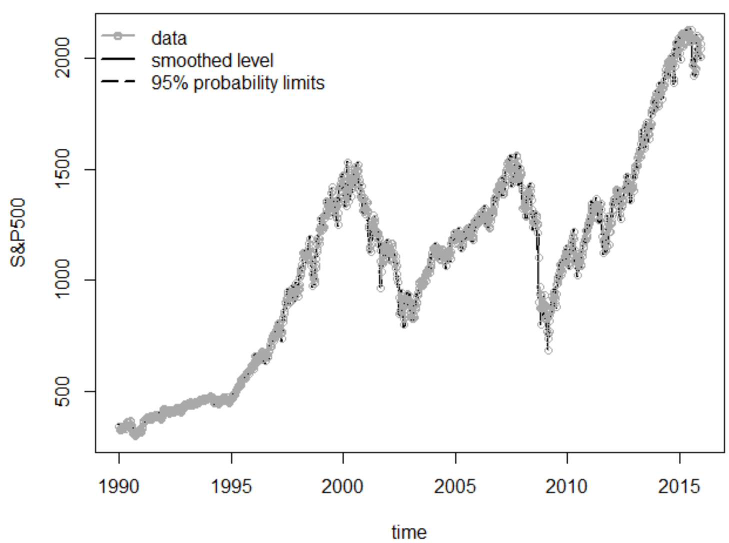
The plot below displays that with Kalman Filter the filtered price resulting from the random walk model. It is apparent that, the filtered values tend to follow the data very closely.



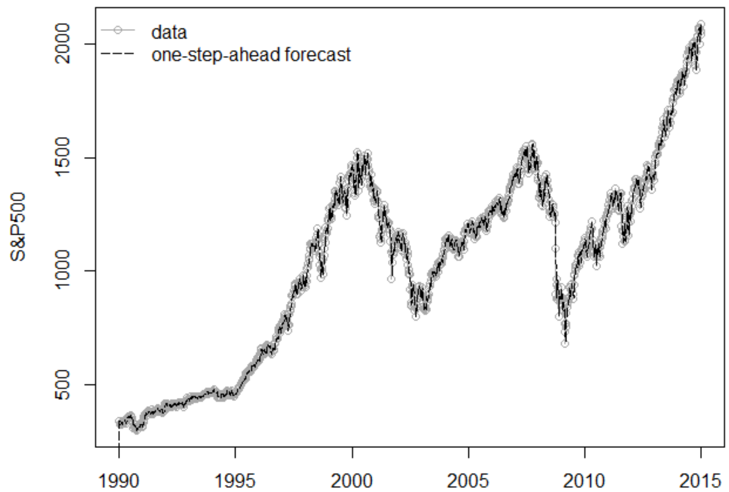
We could see the result more clearly in the following plot. It shows the differences between the Filtered values and the observation data. Here, the range of the price of SP 500 is about 300 to 2000, and the range of the absolute value of the error is less than 20, which shows the Filtered values and the observation data are quite close.



Here is the plot after smoothing, with the 95% probability limits.

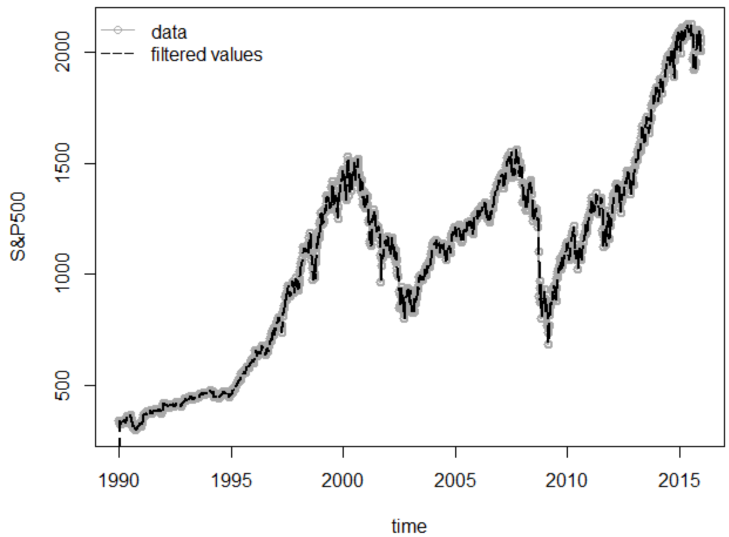


And next is the plot of One- step-ahead Prediction for S&P 500. And we can see that The random walk plus noise model made an accurate one-step-ahead forecast.

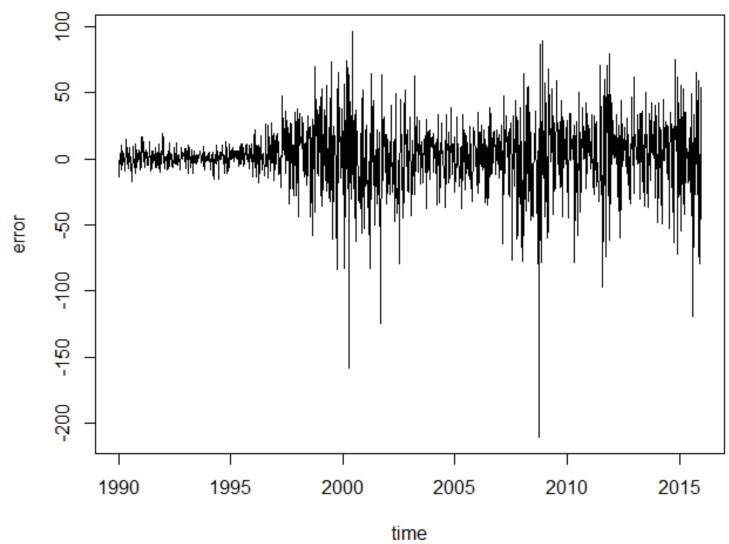


5.2 ARMA Model

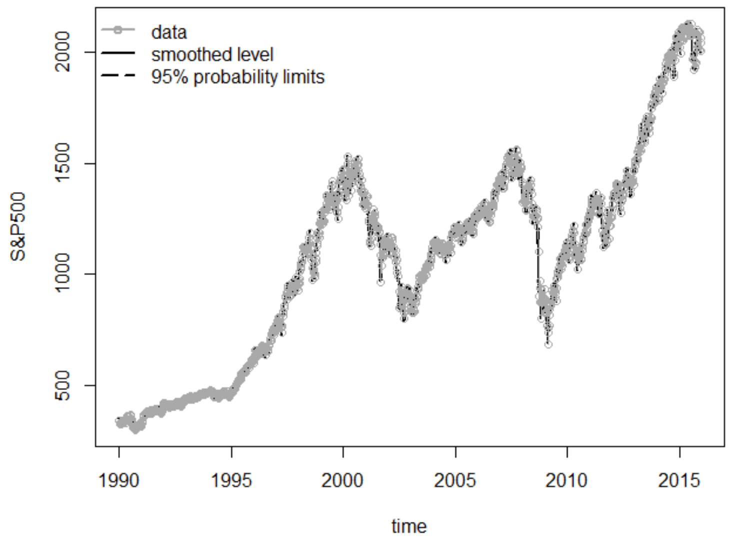
In R we first find the parameters of AR and MA with package "Farma". Then plug it into the "dImModArma" function in the "dlm" package, and then do the Kalman filter. We get the filtered price of S&P 500 Index. It seems that the filtered values tend to follow closely with the data.



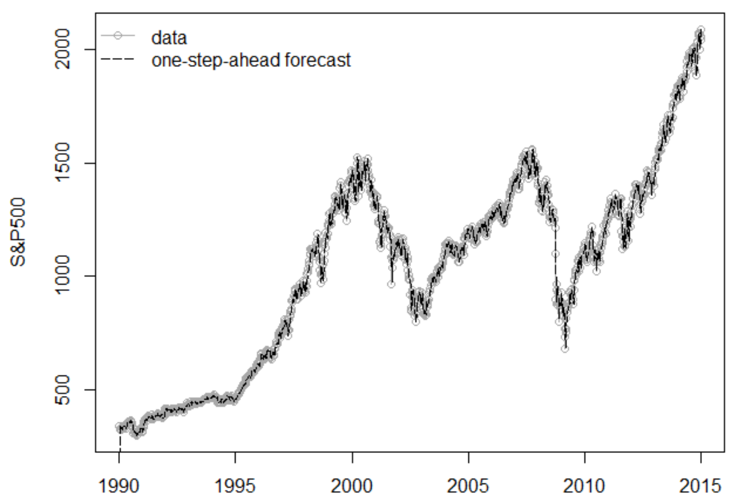
With this plot, we could see more clearly, the absolute value of the difference between the filtered values and the data is less than 200. Which is not as close as the former random walk model.



And then Here is the plot after smoothing, with the 95% probability limits.



And finally, is the one-step-ahead forecast for S&P500. And we can see that The ARMA model in a DLM form also made an accurate forecast, but is not as well as the random walk plus noise model.



6 References

[1] Giovanni Petris, Sonia Petrone, Patrizia Campagnoli, *Dynamic LinearModels with R* ,Springer Dordrecht Heidelberg London New York

[2] Jia Zhou(2009), *Modeling S&P500 STOCK INDEX using ARMA − ASYMMETRIC POWER ARCH models*

[3] E.Scott Mayfield(1992), Bruce Mizrach, *On Determining he Dimension of Real-Time Stock-Price Data*, Journal of Business Economic Statistics.