Financial Time Series Forecasting in a Bayesian Framework

David J. Souza Jr., CFA
P/E Investments
930 Winter Street Suite 2450
781-899-8200
dsouza@peinvestments.com

ABSTRACT

In this paper, we provide an overview of the Bayesian methodology and suggest a tractable online Bayesian framework to forecast financial time series data. The paper is broken up into three sections. Section 1 gives an overview of the Bayesian methodology. Section 2 describes a tractable online Bayesian framework, known as Dynamic Linear Modeling, suitable to forecast financial time series data. Section 3 provides a forecasting example utilizing framework described in section 2.

KEYWORDS

Bayes' Theorem, Bayesian, Dynamic Linear Models, Ordinary Least Squares, Probability

1. INTRODUCTION

Time series forecasting in a Bayesian framework facilitates efficient combination of prior and new information by rules of probability. It is efficient in the sense that all relevant historical information is retained and combined with new information as time moves forward. Forecasting in this way utilizes Bayes' theorem to compute posterior (updated) probability densities from conditional and prior densities when new information is observed. The Bayes' theorem was formulated by 18^{th} Century British mathematician, Thomas Bayes.

1.1 BAYESIAN METHODOLOGY

The Bayesian methodology provides a structured way of routine learning without the need for any particular assumptions [1]. Before each forecast occurs, the forecaster has a prior probability density p(M) describing the uncertainty in model **M**. Model **M** provides a means of forecasting response series **Y** via conditional probability density

$$p(Y|M) \qquad (1.1)$$

and is read "the probability of **Y** given **M**". A useful point to note here is that the forecaster is able to use all historical information known at a given time to forecast a future period.

By the laws of probability, the combination of prior and conditional densities yields a joint probability density namely

$$p(Y_t, M_{t-1}), \text{ or equivalently} \qquad (1.2a)$$

$$p(Y_t|M_{t-1})p(M_{t-1}). \qquad (1.2b)$$

The joint probability density provides a mechanism to recursively revise the model's prior knowledge when new information is available. Specifically, once the forecast period has passed, the modeler will have a new observation, namely **Y***, to update the prior knowledge using the above mentioned joint probability density equation.

The updated (or Posterior) probability density for **M** is proportionately defined by conditional density

$$p(M|Y^*) \propto p(Y^*, M), \text{ or equivalently} \qquad (1.3a)$$

$$p(M|Y^*) \propto p(Y^*|M)p(M). \qquad (1.3b)$$

Equation 1.3b forms the basis for error learning. This updating or learning process about the forecasters' uncertainty in model **M** is often expressed as

Posterior ∝ Observed Likelihood × Prior.

Where the observed likelihood measures the uncertainty in model M based on the new observation.

Bayes' Theorem is defined by dividing the right side of the updating equation 1.3b with the prior density of **Y** or p(Y*). The division simply ensures that the resulting posterior p(M) is normalized to unit probability. The resulting Bayes' Theorem representation is

$$p(M|Y^*) = p(Y^*|M)p(M)/p(Y^*). \qquad (1.4)$$

1.2 INTUITIVE EXAMPLE

A coin toss is a simple example of how a forecast can be revised over time. Before the first flip, prior knowledge would suggest a 50/50 chance of heads or tails. However, after five coin flips all landing on heads, confidence in the prior knowledge would fade. Going forward, the forecaster would put increasingly more weight on heads being flipped next. The coin may have been altered to produce such probabilities. This example is similar to any Bayesian forecast system in pursuit of relationships between dependent and independent variables used for forward-looking projections.

1.3 WHY A BAYESIAN FRAMEWORK?

The Bayesian Framework stands in contrast to the widely-used Ordinary Least Squares (OLS) methodology. Ordinary Least Squares focuses on fitting forecasts to historical data series. Further, OLS tries to fit historical data with the sensitivities assumed constant over the entire data period. This method is not in line with the main goal of forecasting, that is, generating accurate forecasts. By contrast, the Bayesian Framework seeks to minimize forecast error rather than fit historical data. Further, unlike OLS, the Bayesian framework does not assume the sensitivity of one series on another to be constant over time. Rather, a routine, coherent processing of information, with recursive revisions to prior knowledge, guides the future forecasts. The Bayesian Framework accounts for changes in sensitivities over time.

1.4 IMPLEMENTATION OPTIONS

The above-mentioned forecasting and updating process can be computed using the Markov-Chain Monte Carlo (MCMC) methodology to calculate probability densities. However, the MCMC method is computationally intensive and lacks the guarantee of convergence. We propose a tractable online method known as Dynamic Linear Modeling(DLM). DLMs are suitable closed form solutions for forecasting short-term financial time series data.

2. DYNAMIC LINEAR MODEL

DLMs provide a two-stage tractable online Bayesian framework to forecast financial time series data. The two stages are forecasting and updating. The model form is defined as follows for a one factor forecasting system [1].

Observation equation: $Y_t = F_t\theta_t + \nu_t, \quad \nu_t \sim N[0,V_t]$ (2.1a),
System equation: $\theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0,W_t]$ (2.1b),
Initial information: $(\theta_0|D_0) \sim N[m_0,C_0]$ (2.1c),

where:

- Y is the response vector to be forecast;
- F is a known factor or regressor variable vector;
- θ is the state, sensitivity, or beta vector;
- G is the evolution or state matrix;
- D is data known at specified time;
- ν is the observation error;
- ω is the system equation error;
- V is the observation equation variance;
- W is the system equation variance;
- m is the prior mean beta;
- C is the prior mean beta variance.

Response series **Y** (asset return) and **F** (factor level) provide information needed to learn about the **θ** (Beta relationship) between each other and ultimately to forecast future values of **Y**.

2.1 OBSERVATION EQUATION

The observation equation for the DLM defines the sampling distribution for $\mathbf{Y_t}$ conditional on $\boldsymbol{\theta_t}$. This equation relates $\mathbf{Y_t}$ and $\boldsymbol{\theta_t}$ with a goal of specifying observational variance $\mathbf{V_t}$. The conditional structure is

$$(Y_t|\theta_t) \sim N[F_t\theta_t, V_t]. \tag{2.2}$$

$\mathbf{V_t}$ is a valuable component in the correction term used to update $\boldsymbol{\theta_t}$. $\mathbf{V_t}$ is a measure the inverse precision in the observation equation. All else held constant, when $\mathbf{V_t}$ is very large, the amount of adaptation is low: by contrast, when $\mathbf{V_t}$ is very small, the amount of adaptation is high.

2.2 SYSTEM EQUATION

The system equation defines the evolution of the state vector. The conditional structure is

$$(\theta_t |\theta_{t-1}) \sim N[G_t\theta_{t-1}, W_t]. \tag{2.3}$$

W is the system equation variance. Specification and magnitude of **W** is crucial to the model learning as it defines the stochastic evolution and stability. In other words, **W** determines the durability of the model. If **W** is low, the reliability of the model is high. However, the model becomes increasingly unreliable as **W** grows. Yet another way to think about **W** is as a measure for adapting with new information. When **W** is low(high), model is reliable(unreliable) and adaptation of new information is low(high).

2.3 TWO-STAGE PROCESS

The two-stage process entails forecasting and updating. Processing begins once the response and the regressor series are specified inside the DLM structure and the initial prior θ_0 mean and variance have been obtained. Generally, the initial prior mean and variance states rely on historical data, or reflect educated guesses on the part of the forecasters. In either case, the initial prior information should be quite diffuse and have minimal impact on the current forecast as time moves forward. In other words, the DLM will learn and revise the forecast dramatically in the early periods if the initial priors are significantly off the mark. In addition to obtaining priors to begin the two-stage process, care must be taken to ensure the correct lagging of factor data to avoid using hindsight in the forecast.

With the DLM specified and priors in hand, the system provides a means of forecasting $\mathbf{Y_t}$. This is the first stage of the process. The second stage provides a means of sequential revision to the prior knowledge when new observations become available. The new observation is compared with the forecast to quantify the amount of forecast error. Forecast error is the main driver of level changes in the system variance **W** and as result system learning. A large forecast error acts as a red flag to the model. If large forecast errors persist, the DLM will become less confident about the prior knowledge, the system variance **W** will rise and increase the amount of adaptation needed to learn the new relationship. Said differently, the revised, also called posterior, Beta at each time step is obtained by correcting the prior Beta with a term proportional to the current forecast error. The revised prior becomes posterior knowledge, which is used as prior information for the next forecast period. The above process continues indefinitely. To recap, the two-stage process is as follows:

> ***Forecast Stage***- Prior information is used to forecast one-step ahead predicted **Y** level and the variance.
>
> ***Update Stage*** - Observation is taken and compared with the forecast. A term proportional to the current forecast error defines the amount of updating or learning needed to revise prior knowledge. The revised prior is now called the posterior, which is the prior information for the next time period.

This two-stage forecasting and learning system is continued until all data has been processed. Once again, Bayesian learning is best illustrated by the following equation:

$$\textbf{Posterior} \propto \textbf{Observation Likelihood} \times \textbf{Prior}$$

Figure 2.1

Such a dynamic model follows a Markovian stochastic evolution of the quantity under investigation. In a factor DLM, the Beta relationship $\boldsymbol{\theta}$ would follow the Markovian structure as illustrated by the following influence diagram.

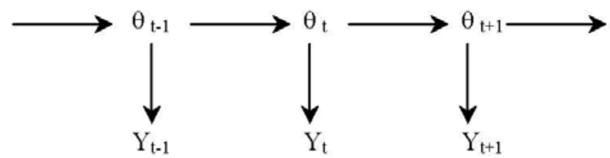

Figure 2.2

3. FORECASTING EXAMPLE

The following forecasting example employs the factor DLM framework described in Section 2. This example applies the DLM framework to forecast excess returns in developed market currencies using short-term interest rate differentials. Here, the quantity under investigation is the Beta relationship $\boldsymbol{\theta}$ between developed market currencies and short-term interest rate differentials (STRD). STRD is well known as the "carry trade" strategy, where a speculator will buy high yielding currencies and short low yielding currencies.

The chosen example forecast period, Q3 2006 through Q2 2008, is a period during which many "carry" only currency strategies failed to make significant positive excess returns. This raises two questions regarding the DLMs forecasting ability. First, did the system adapt during this period? Second, how did the DLM perform versus some well known "carry" benchmarks [2] during this period?

The first question can be answered by looking at a graph of system variance **W** over time. Recall from section 2, large forecast errors show up as higher levels of **W**. A higher **W** leads to increased system learning. Figure 3.1 below illustrates just this. Between September 2006 and December 2007, the level of system variance rose continuously as forecast errors persisted.

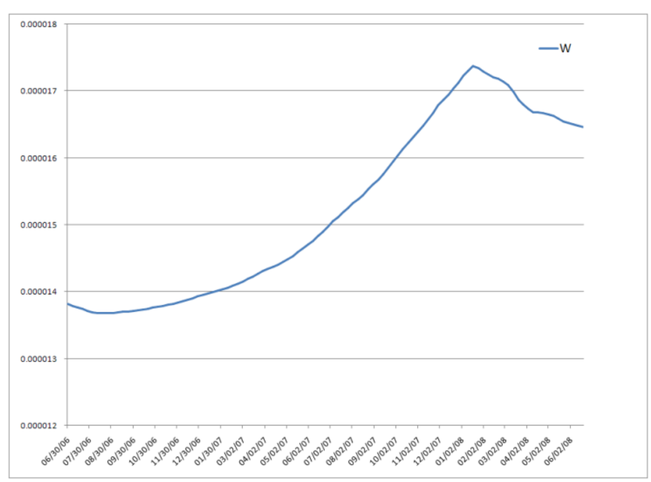

Figure 3.1

With the DLM system learning from forecast errors, attention is naturally drawn to the second question raised above. How did the system perform? Figure 3.2 illustrates the value of a $1000 investment in the factor DLM(green), as well as three "carry" indices, the Deutsche Bank Carry Index(orange), the Barclays Carry Index(gray), and the Credit Suisse Carry Index(turquoise), from July 1, 2006 to June 30, 2008. The DLM exhibited superior performance going into the "carry trade" performance peak in the summer of 2007, as well as during the subsequent draw-down period.

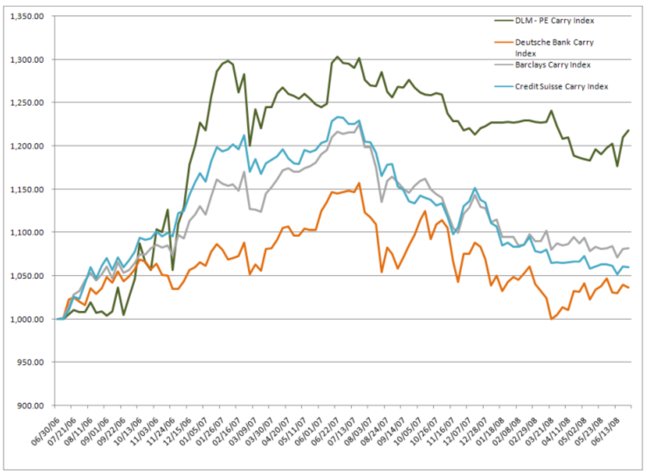

Figure 3.2 (source: P/E Investments and Bloomberg)

On a risk adjusted basis, Table 3.1 shows an attractive information ratio for the factor DLM relative to the information ratios achieved by the benchmark strategies.

Strategy	Average Ann. Return	Ann. Standard Deviation	Information Ratio
DLM - P/E Carry Index	10.4%	12.5%	0.83
Deutsche Bank Carry Index	1.8%	9.9%	0.18
Barclays Carry Index	4.0%	7.8%	0.52
Credit Suisse Carry Index	2.9%	7.8%	0.38

Table 3.1 (source: P/E Investments and Bloomberg)

4. CONCLUSION

This paper has provided an overview of the Bayesian methodology in section 1 and the DLM in section 2. Financial time series forecasting within such a framework may prove superior versus other methodologies as shown in section 3.

5. REFERENCES

[1] West and Harrison 1997 "Bayesian Forecasting and Dynamic Models" 2nd edition.

[2] Bloomberg - Deutsche Bank Carry Index (ticker FXCARRSP), Barclays Carry Index (ticker BXIICRUS), Credit Suisse Carry Index (ticker ROCI10US)