Article can be accessed online at http://www.publishingindia.com

Airline Revenue Management – Revenue Maximization Through Corporate Channel

Karthik V*, Indranil Mitra*

Abstract

This paper explains an empirical model that has been developed to arrive at the Best Possible Fare (BPF) for all the ad-hoc requests made by the corporate passengers. For the purpose of this research, a corporate request is considered ad-hoc if the booking request is initiated after the corporate channel for the particular flight is closed. By charging the best possible fare, the airline will be able to marginally increase its revenue without deviating from the guidelines of the corporate channel. This model updates itself with the available capacity at the time when the ad-hoc request is initiated, also considers the previous booking data to forecast the passenger demand and the channel behavior. This will lessen the manual intervention and its associated errors, and will take care of the number of corporate requests that can be approved and size of the corporate booking requests that can be approved. As the factors affecting the booking trend of the airlines have been covered earlier in various research papers as discussed in the literature review, we have directly focused on deriving the empirical solution in this paper.

Keywords: Revenue Management, Best Possible Fare, Forecasting, Airlines, Corporate Passengers, Sales Channel

INTRODUCTION

After the deregulation of airline industry in India, all the airline organizations started offering differential price for the seats in the same flight and in the same cabin. Different airlines use different pricing rules and algorithms to arrive at the final price for each seat which is to be sold at a future date. With the invasion of different platforms in the form of travel agents and online booking systems to compare and book tickets, the Indian Airline Industry is suffering to maximize its revenue in the low cost game. This part of how many seats should be offered at what rate and for which period of time before booking for a flight closes, is taken care by revenue management team. The importance of revenue management to dynamically vary their fares according to the supply and demand is essential in the path towards maximizing the overall revenue. Each and every airline manages to fulfill this objective depending on the level of sophistication of the algorithms and scientific techniques incorporated into the system.

Typically, historic booking data, information of seats sold through different channels and its forecasts are used to predict the future demand. This forecast serves as an input for the revenue optimization step, which considers capacity and fares (Bamberger, G.E., D.W. Carlton and L.R. Neumann., 2004). The resulting inventory controls are balanced with real world demand when taking reservations. Demand is strongly influenced by market conditions. The results obtained are again used as input data for the next demand forecast and so on.

In airline industry the customers can be broadly divided into categories; corporate and leisure (Kelly Mc. Guire, 2012) based on their ticket booking behavior. As the need to travel on a particular date is different for each of these two categories, the airline uses differential pricing methodology to extract maximum value out of the customer to enhance its revenues. The price of the ticket for a leisure passenger increases as the booking date comes closer to the departure date. This is because, these passengers who look for tickets during that time have higher propensity to travel and hence will be ready to pay the premium to purchase the ticket. As these customers are individual customers and they are not bound by any contract, the airline does not see any benefit in giving them a discount on the maximum value that can be extracted from the customer. However the price of the ticket for corporate passenger does not vary much because of the volumes of the business the corporate brings to the airlines.

* KPMG-India, D&A-Management Consulting Gurgaon, Haryana, India.
E-mail:karthikv606@gmail.com, Indranil.mitra@gmail.com

In the case of corporate passengers, the airline enters into a contract with the corporate customer to provide special benefits with less degree of variation in the ticket price without any effect on the date of travel. Due to this reason the airline is not able to alter its fare in case of a special request from an existing corporate customer leading to opportunity loss in terms of selling tickets at increased price through other channels.

The change in the pricing structure is governed by various pricing rules and algorithms written for the same. It is observed that in general the pricing algorithm is a factor of demand, competition pressure, sales channel, and booking time as a function of days before departure among others (Fedorco & Hospodka, 2013).

This paper is laid out as follows: In section 2, we have illustrated the current scenario as the problem definition and emphasized on the scope for improvement in the same section. To build an objective case we have also made a few assumptions which are also mentioned in the same section. In section 3, we have reviewed the contribution of a few research papers in exploring how dynamic pricing across different channels has been handled in the past. In section 4, we have explained about our model and the approach considered in building the formula. In section 5, we have developed the empirical formula that will give us the best possible fare that should be charged to accept the special request from the corporate customer. In section 6, we have discussed the results and consolidated the investigations. In section 7, conclusion of the paper is presented.

PROBLEM DEFINITION

In Industry various revenue management solutions are available in the market to help airlines arrive at the optimal passenger mix for a scheduled aircraft. But these solutions are built to recommend the optimal passenger mix of corporate and leisure passengers at a suitable time before the seats are booked for a particular flight. Based on the recommended passenger mix, the airline opens its inventory for both classes of passengers across various channels. Once the aircraft reaches its recommended passenger mix for a particular type of passenger, the inventory for that particular type of passenger is blocked and the remaining seats are allotted for the other passengers.

Consider this scenario: The corporate bookings for a particular flight with departure date 7 days from now are closed, as the revenue through corporate channel has reached the desired limit. Now there is a special request of size 'S' for additional bookings from one of the corporates for the same flight. The revenue management team has to decide whether to accept the request or not.

Under the current scenario, the airline

1. Accepts the request, if it can achieve the budgeted flight revenue.
2. Accepts the request, if the corporate is a loyal customer.
3. Rejects the request otherwise.

In each of these conditions, the decision is arrived at by performing basic projection of the commercials based on past performance and future relationships. The airline does not use any scientific reasoning to forecast the opportunities across other existing channels. A standard revenue management solution does not have the provision to accommodate the ad-hoc requests. The requests, if approved, are approved at the previously determined rates. Hence the airline is not able to leverage the left out inventory to make additional revenues through more profitable channels.

To approve 'S' seats for corporate ad-hoc requests, the airline displaces 'S' seats from some other channel. Since the price of seats in other channels are highly sensitive to the date of travel, while borrowing seats from other channels the airlines must consider the opportunity cost and arrive at a suitable price to avoid losses. The corporate ad-hoc request can be accepted as long as the revenue generated from this scenario is at least equal to the expected revenue that the corporate passenger displaces.

In the scope of the paper we do not consider the external factors that impact the booking trend before the departure of the flight. External factors like natural calamity, any unprecedented event in the arrival or departure destination and others. We have predominantly focused on the industry trend and controllable seasonal factors, which under normal scenario do not alter the booking trend of a flight.

Assumptions: To make this analysis quantifiable, we have made the following assumptions in the paper.

1. When a passenger considers to travel, he/she approaches the airline mentioned in the corporate agreement.

2. We have considered corporate passengers only through corporate channel as it difficult for us to differentiate between passengers through other channels.
3. We have considered customers as if purchasing one way tickets only. Since there is less difference between the price of one way and round trip tickets a customer who has purchased a round trip ticket is considered to have purchased two one way tickets.
4. We have assumed airlines utilize finite set of fares. Though the airline sells the inventory at different prices across different channels, airlines chooses a single set of fare to offer at a particular period of time.
5. Cost centers are fixed. Since the marginal increase in the revenue does not come at the cost of any operational expenses, increase in revenue leads to increase in profitability.
6. Finally, we assumed that airlines do not oversell tickets as most of the routes studied are operated by airline that do not oversell tickets.

LITERATURE REVIEW

The question of identifying the optimum passenger mix and setting the price for each channel and time period has been discussed a lot in the academic literature.The considerations that started much of revenue management research can be found in Littlewood (1972). He mathematically formulated an intuitive rule that proposes to sell tickets in the cheaper of two booking classes, as long as the expected marginal utility exceeds the fare of the more expensive booking class. Later, Belobaba (1987) extended this approach to more than two booking classes, resulting in a still frequently applied concept: expected marginal seat revenue (EMSR). A review of developments and future directions of revenue management are given by McGill and van Ryzin (1999). Furthermore, Talluri and van Ryzin (2004) provide a comprehensive insight into the concept of revenue management and its elements. A more recent overview of mathematical models and methods in revenue management and its focus on simulations can be found in Talluri et al. (2008).

There are two important streams of research on revenue management: (i) empirical studies on airline pricing methodologies, and (ii) analytical revenue management models in the literature. In the economics literatures, studies have empirically examined the relationship between airline pricing and various market factors. Borenstein and Rose (1994) find a significant positive effect of competition on price variations in the airline industry. Hayes and Ross (1998) find airlines' price discrimination policies lead to increased price variations. Borenstein (1989, 1990) finds that airport dominance enhances a carrier's ability toattract passengers and charge higher fares. This may be attributed to biases due to computerreservation systems, the dominant carrier's local reputation, control of critical inputs such asgates and slots, and marketing strategies such as frequent flier plans (Evans and Kessides 1993).Peteraf and Reed (1994) find that a monopolist's national market share has a positive effect onfares and that prices tend to decrease in the number of passengers and route distance.

Moreover, they find that the average code-share fare is lower than the average fare that isnot code-shared. Bamberger et al. (2004) also find that the price tends to decrease after alliances.Their findings are similar to those of Park and Zhang (2000), Brueckner and Whalen (2000), andBrueckner (2001, 2003) who examined international alliances.

The quantity-based revenue management models start with Littlewood's seminal work (Littlewood 1972, henceforth referred to as the Littlewood model). The Littlewood model studies how the fixedtotal capacity should be allocated between two classes of seats once fares are determined. The model assumes a fixed number of seats and two independent classes of demand-demand forfull-fare tickets and demand for discount-fare tickets. Discount-fare demand occurs first, and it is large enough to fill all the allocated seats. The demand for full-fare tickets occurs later and is random. The model derives the optimal seat protection level for full-fare demand. The analysis of the problem is similar to that of the classical newsvendor problem in the inventory theory (Talluri and van Ryzin 2004). The Littlewood model has since been extended to multiple-classmodels (Belobaba 1989, Wollmer 1989, Curry 1990, Brumelle and McGill 1993, Robinson 1995) and dynamic models (Lee and Hersh 1993, Feng and Xiao 2001).

For price-based revenue management models, the seminal work of Gallego and van Ryzin (1994) analyzes the optimal dynamic pricing policy for one type of product. Gallego and van Ryzin's dynamic pricing model assumes that consumers arrive randomly. The optimal price has the following important properties: (i) At any fixed point intime, the optimal price decreases in the inventory level; conversely, for a given level of inventory level, the optimal price increases with more time to sell. (ii) For a fixed time and inventory level,the optimal price increases in the arrival rate. Zhao and Zheng (2000) extended this model to thecase where demand is non-homogeneous. Since consumers are time sensitive, their reservation price distribution may change over time. For a good review of the current practices in dynamic pricing, see Elmaghra by and Keskinocak (2003).

The Littlewood and GVR models offer important insights that will be used in our discussion of the empirical findings. Shumsky (2006) finds that low-cost competitors are driving the network airlines to rely on alliances for an increasing proportion of their traffic. Wright et al. (2006) study a variety of static and dynamic mechanisms to manage revenue management decisions across alliances.

Using these mechanisms the airline is able to change its pricing for the same ticket based on the channel and the time period of search of the ticket. Airlines tend to charge high prices from passengers who search for tickets close to the date of travel. The conventional view is that the airlines capture their high willingness to pay through inter-temporal price discrimination. Airlines also adjust price on a day-to-day basis as capacity is limited and the future demand for any given flight is uncertain. While fares for a leisure passenger generally increases as the departure date approaches, fares for the business passengers are not altered much because of the corporate agreements and the quantum of business it generates during the engagement. By keeping the ticket price constant irrespective of the time period of ticket being booked and the type of corporate the airlines are losing out on the opportunity to marginally increase the revenue. Hence in this research paper, we have considered the factors impacting the sale of tickets through corporate channel and arrived at the best possible fare that should be charged from a corporate to maximize the revenue without breaching the guidelines of the channel.

Our Work

Based on the ticket booking behavior all the airlines have divided their target customers into leisure and corporate passengers at a broad level as mentioned in the literature review. Each airline has its own strategy to tap its target audience i.e. channel, pricing and services offered for the same seat in a particular flight.

Most of the airlines today manage corporate bookings through their specific corporate booking channel and also through specialized indirect agents who help in acquiring only corporate customers. The entire process of acquiring a corporate customer is bound by a contract between the airline and the corporate customer, hence there is less manual intervention at the time of booking the ticket. This leads to less flexibility in associating the latest available price of the ticket to the corporate customers. Since major part of the corporate bookings happen less than one week before the date of travel, the price of a ticket during that time through corporate channel is less than the cost of the same ticket through other non-corporate channels.

In the case of a flight where the corporate bookings are closed after the channel reaches its desired load factor any additional request is separately handled by the revenue management team. The entire process involves manual intervention from the airline side – right from enquiry for additional corporate booking through approval, booking, modifications to the itinerary, payment and ticketing. Entire process is resource intensive, time consuming and error prone right from enquiry to final billing. Also if the corporate special request gets approved the ticket is booked at a predefined pricing rule as per the contract. It is generally observed that the same ticket through other channels is higher than what is actually sold. Hence the airline has lost the opportunity to earn additional revenue by selling the ticket at a lower price through corporate channel. The difference in the pricing through these channels is so high that the airline ends up selling the ticket at a price less than what it could have sold by charging a premium without comprising the benefits offered to corporate passengers for the bookings done during the same time. The problem is further complicated by last minute cancellations and modifications on PNR resulting in inventory being blocked which otherwise can be sold through other channels and then maximize the revenue quotient for the airlines.

Keeping the above factors in mind we have arrived at an empirical solution from a representational sample of three months booking data of a leading airline organization. This empirical solution is helpful in providing decision support through scientific forecasting and optimization techniques to arrive at the best possible fare that needs to be charged to accept the corporate special request below which it does not make business sense to accept the corporate special request. This will lead to marginal increase in revenue by charging a premium to approve their special requests without deviating from the agreement. This is done with respect to analyzing the historical booking data at channel levels, corporate booking activities, utilization rates of corporate bookings and many more.

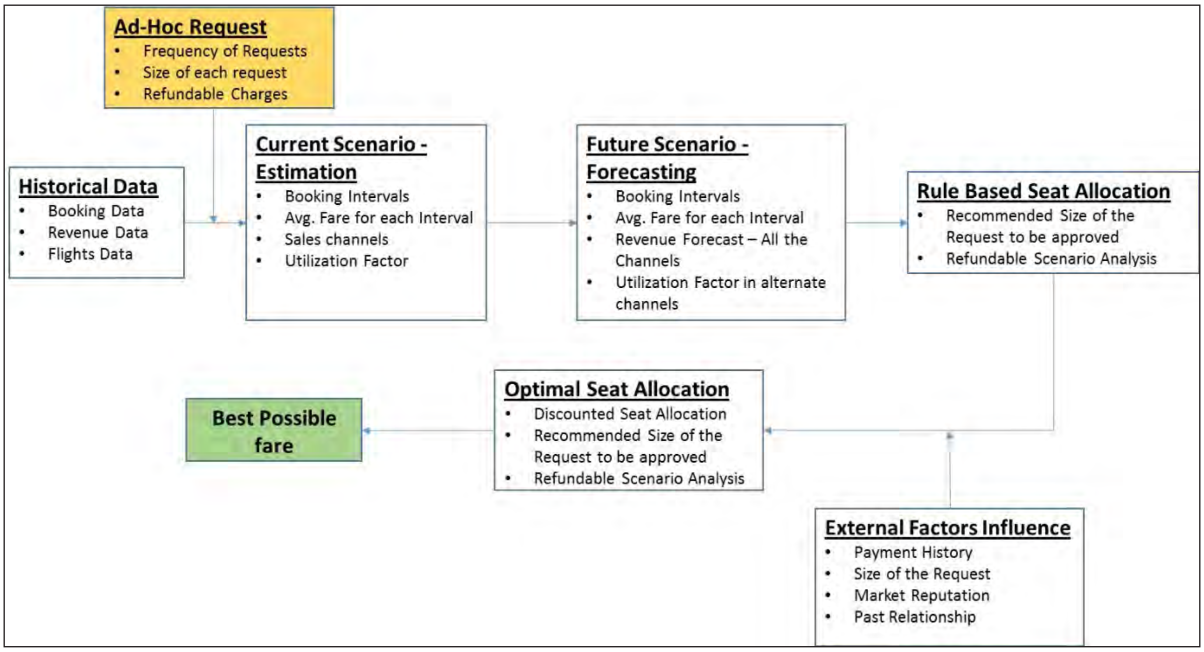
METHODOLOGY AND FRAMEWORK:

Most airlines today manage their corporate booking through their specific corporate tie ups in the form of direct agreements with a corporate or indirectly through specialized corporate agents who are directly associated with the corporate. The decision of accepting the request, pricing of tickets, ensuring the arrived price is the maximum value that can be charged for the booking without losing the customer, special requests like internal travel can be complex based on the requests and cancellations. The entire process involves manual intervention from the airline revenue department rendering the whole process tedious and error prone.

The figure below gives a schematic layout of the approach adopted to calculate the best possible fare.

This methodology

1. Provides an optimal airline seat allocation scheme for the revenue management team for their additional corporate requests
   a. Number of seats for corporate booking process
   b. Number of corporate bookings request
   c. Size of each booking request
   d. Corporate Price Per Seat
   e. Revenue Forecast
   f. Refunds in case of cancellation
2. Provides an optimal control mechanism for the corporate booking process
   a. Demand Pattern
   b. Booking Pattern
   c. Cancellation Pattern
   d. "No Show" Pattern'
   e. Overbooking
   f. Corporate booking Start Date
   g. Corporate Booking End Date



   h. Time between negotiation phase and placement of non-refundable deposits
   i. Time between non-refundable deposits and actual price of tickets
   j. Time between actual purchase of tickets and date of departure

Thus, there are several other factors that need to be incorporated in a structured manner to come to a reasonably good solution.

A random sample of the past three month's data was extracted for all the considered routes (metro). A total of 35 variables were chosen, of which through correlation analysis, 9 variables with higher degree of correlation were dropped. Remaining 26 variables are considered for the study. Subsequently we performed ANOVA test to identify the degree of significance of each variable. Since 8 variables had $p>0.05$ we have finally kept only 18 variables for our study. The Cronbach alpha of the chosen set of variables is found to be 0.6723 which is within acceptable limits for the study. The list of the variables used for ANOVA test and their p value are presented in Appendix 1.

To consider only flights where the number of ad-hoc requests received are consistently more than one we have considered all the flights with load factor > 80%. In terms of number of flights this data forms 40% of the entire dataset considered.

To arrive at the best possible fare we need to calculate the Total Expected Corporate (TECR) revenue from the corporate special request to make the flight profitable. TECR is calculated by analyzing all the possible scenarios of seat allocation by taking booking intervals, utilization factor for each channel among others into consideration. The Next step is to calculate the Total Expected Flight Revenue (TER) for the same number of seats displaced from other channels. In this step, we will be using statistical forecasting techniques to calculate the forecasted revenue across all the channels. If the difference between TECR and TER to fill the same number of displaced seats under normal scenario is positive then calculate the best possible fare by dividing the TECR by size of the request. If the obtained average fare is less than the average fare as mentioned in the agreement then the revenue team quotes the price as mentioned in the agreement.

MODEL

Corporate booking process starts about 12 months prior to the flight departure. Let the time span from the start of the booking till flight date be divided into suitable intervals as T = t, t-1, t-2,…, 1, 0. Where t=0 denotes the actual flight departure time. For each time point we have the available seats, demand for the available seats across each channel, average fare for the seats and the actual request that comes for the corporate booking. To illustrate further consider the following:

Let,

The set be defined as:

I – Class of fares for an airline

K – Total Time Interval starting from the start of corporate booking till departure of flight

Let the variables be defined as,

At each point there are 3 components to corporate booking process:

D_{ij} – Demand that is estimated from the historical data at the time point 'j' for the 'I' class

F_{ij} – Average Fare that is to be estimated from the historical data at time point 'j' for class 'I'

μ_{ijr} – the corporate request that comes at the time point 'j' for the class 'I'

C_i – the number of reserved seats for corporate booking for the class 'I'

G – Total number of booking class

$$\sum_{i=1}^{g} Ci = C,$$

F_{ik} – Average Fare for the i^{th} class at a particular time interval 'K'

C – Corporate booking limit for the i^{th} class

– Opportunity cost of the airlines because of not selling the seats through some other class of service

– Fraction of originally booked tickets at k^{th} point of time.

– Opportunity cost of the airlines because of passengers who paid for the seat but did not show up

– Fraction of originally booked tickets at k^{th} point of time.

Decision Variable

– The number of seats that will be allotted to a corporate request (r) of size at time point 'k'.

Here we are assuming that a corporate request of μ at a particular time point comes for the i^{th} class.

Mathematical Model

The base model that would be used for optimization can be given as:

$$z = Max \sum_{r} \sum_{k} \{ (\sum_{i=1}^{g} F_{ik} X_{ik}) - \left(C_{1kr} * P_{ikr}^{c} + C_{2kr} * P_{ikr}^{c} \right) \} \boxminus (P)$$

Subject to

$$X_{ikr} <= \min (d_{ij}, \mu_{ikr}) \quad (1)$$

$$\sum X_{jkr} <= C \quad (2)$$

At any point 'K', gives the revenue earned over all the classes for a particular flight. On the other hand ($C_{1k}*P^{C}_{1k}+C_{2k}*P^{C}_{2k}$) gives the total opportunity cost that the airline incurs at the time point 'K' taking into consideration the different situations as described before. Thus subtracting the opportunity cost from the revenue earned gives net revenue earned at time point 'K'.

Constraint (1) implies that the number of seats offered would be less than or equal to the number for which order is placed.

Constraint (2) implies that the number of seats offered in all classes is to exceed the capacity offered in corporate request.

Model output of P gives the size of the corporate. Solving the objective function of (P) gives the maximum revenue that can be earned at k^{th} point of time where k = K.

Best Possible Fare Calculation

In accepting a corporate request of size 'S', an airline potentially displaces up to 'S' individual passengers. Since corporate fares are discounted below the fares through other channels as the booking date gets closer to the departure date, the decision whether or not to accept a corporate request depends on individual passenger on each flight flying in comparison with corporate passenger. This corporate request should be accepted as long as it makes business sense to the airlines. This is termed as the total expected revenue of the displaced passengers(TERDP).

Best Possible Fare (BPF) = TERDP/No. of approved Corporate Request

Where TERDP is calculated from the historic booking trend of the same aircraft for a similar booking season.

Let us define Z(C) to be the optimal objective value function (I) using the initial set of corporate booking limit C.

Now consider an ad-hoc request of size 'S'.

Z(C-S) – is the optimal objective function solving (I) where the capacity constraints of each aircraft where the corporate passenger will travel is decreased by the size request S.

The value Z(C-S) is the best solution of the problem given that one has accepted the corporate request and S seats are no longer available for further passenger booking.

We define the difference of the objective functions Z(C) and Z(C-S) to be D(S) which is defined as TERDP. Thus D(S) represents total expected revenue of displaced passengers.

In algorithmic form the following can be proposed:

Step 1: Find Z(C) using the linear mathematical programming formulation (P) for the given network.

Step 2: Reformulate the mathematical program to calculate Z(C-S), where the capacity constraints used in step 1 reduced by S where the group travels. The reformulated model isgiven as:

$$Z = Max\,?^{g}_{(i=1)}??(F_i k X_i k) - (C_1 k P_1 k^c + C_2 k P_2 k^c\} \text{ (Q)}$$

Subject to

$$X_{ik} <= \mu_{ik} \quad (1)$$

$$\sum X_i + S <= C \quad (2)$$

k = K, i = I;

Note here that the constraint (2) states that the capacity of the seats that has been reduced by 'S'

Step 3: Find D(S) = Z(C) – Z(C-S)

Step 4: MAF = D(S)/S

The steps are to be executed in a looping procedure where the next starting point would beZ(C-S). The procedure would continue till all the seats are used up or group request period ceases, whichever may be earlier.

RESULTS

Airlines manage corporate ad-hoc requests based on the potential of business from the corporate. The entire process involves manual intervention from airline revenue department rendering the whole process tedious and error prone.

Using this empirical solution the airline will be able to know the best possible fare below which, by approving the corporate request the airline is losing the opportunity to earn additional revenue. Making an informed decision will also help the airlines plan its future inventory across various channels and adjust its targets to achieve overall profitability. We have summarized below the results that can be obtained by using this scientific way of addressing the corporate ad-hoc request:

S.No.	*Factors*	*Description*	*Drivers*
1	Determine the Best Possible Fare(BPF)	BPF for a given O&D, Date	Displacement Cost Calculation, Revenue Optimization and Inventory Management
		Extend BPF calculations to multi leg journey	
		Self-adjustable BPF due to nearing departure time, cancellations and increased demand	
2	Calculate Discount Rates to be given	Determine discount percentage to be applied on fares based on past performance	Frequent Flyers, volumes of business by the corporate
3	Alternative Travel Plan	Requests that cannot be accommodated for can be re-routed for an alternate date	Network Plan along with Booking Details, Displacement Cost

CONCLUSION

We have used the representative data of a leading Indian airline to examine the standard revenue management practices and perform our analysis. We have concentrated on the revenue opportunities through corporate channel. Because our model includes market factors affecting the revenue of the channel, our empirical formula can also be extended to model other channels.

This paper explains the empirical model to arrive at the best possible fare, required to increase the revenue and hence the profitability without altering the benefits offered to the corporate customers. This will help the airlines to marginally increase their revenue for the same services offered to the customer based on the time of the request, size of the request and passenger mix of the particular aircraft. This analysis is not helpful in a case where the airline feels it is necessary to accept the request as a directive from a higher official without thinking about the impact of accepting the request on its revenue and profitability. Since the number of such important requests are going to be very less as such decisions will have a negative impact on revenue, we have ignored such scenario from our analysis.

Our model helps the revenue management team to arrive at the best possible fare for the corporate ad-hoc request. The current work has the limitation that, in case there are empty seats closer to the date of departure, the model can be extended to come up with discounted fares for high performing corporates as a reward for their loyalty.

References

Alderighi, M., Nicolini, M., & Piga, C. (2012). *Combined effects of load factors and booking time on fares: Insights from the yield management of a low-cost airline*. Milan, Italy: Nota di Lavoro, Fondazione Eni Enrico Mattei.

Bamberger, G. E., Carlton, D.W., & Neumann, L. R. (2004). An empirical investigation of the Competitive effects of domestic airline alliances. *Journal of Law and Economics*, 47, 195-222.

Baltagi, B., Bresson, G., & Pirotte, A. (2003). Fixed effects, random effects or Hausman-Taylor?A pretest-estimator. *Economics Letters*, 79(3) 361-369.

Belobaba, P. (1989). Application of a probabilistic decision model to airline seat inventory control. *Operations Research,* 37, 183-197.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, New York, N.Y.

Bergstrom,T., & MacKie-Mason, J. (1991). Some simple analytics of peak-load pricing. *Rand Journal of Economics,* 22, 241-249.

Borenstein, S. (1989). Hubs and High Fares: Dominance and Market Power in the U.S. *Rand Journal of Economics*, 20(3), 344-365.

Borenstein, S. (1991). The Dominant-firm advantage in Multiproduct Industries: Evidence fromthe U.S. Airlines. *The Quarterly Journal of Economics,*106(4), 1237-1266.

Borenstein, S., & Rose, N. L. (1994). Competition and Price Dispersion in the US Airline Industry. *The Journal of Political Economy*, 102(4), 653-683.

Brueckner, J., & Whalen, T. (2000). The price effects of international airline alliances. *Journal ofLaw and Economics,*43, 503-545.

Brumelle, S. L., & McGill, J. I. (1993). Airline seat allocation with multiple nested classes.*Operations Research*41, 127-137.

Burdett, K., & Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51(4), 955-969.

Cachon, G., & Terwiesch, C. (2006). *Matching supply with demand.* McGraw-Hill/Irwin.

Clark, R., & Vincent, N. (2012).Capacity-contingent pricing and competition in the airline industry. *Journal of Air Transport Management*, 24, 7-11.

Carpenter, D. (2007). *Fuller planes pay off for UAL bottom line.* The Grand Rapids Press, July 25.

Curry, R. E. (1990). Optimum seat allocation with fare classes nested by origins and destinations. *Transportation Science,* 24, 193-203.

Dresner, M., & Windle, R. (1992). Airport dominance and yields in the U.S. airline industry. *Logistics and Transportation Review*, 28(4), 319-339.

Eblen, T. (1996). The grounding of Value jet Airlines must try to fill all seats, charge the highest possible fare. *The Atlanta Journal-Constitution,* June 23.

Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: research overview, current practices, and future directions. *Management Science,* 49(10), 1287-1309.

Evans, W., & Kessides, I. (1993). Localized market power in the U.S. airline industry. *The Review of Economics and Statistics,* 75, 66-75.

Gallego, G., & van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochasticdemand over finite horizon. *Management Science,*40, 999-1020.

Greene, W.H. (2003). Econometric Analysis.Prentice Hall, Upper Saddle River, N.J.Hayes, K. and L.B. Ross. 1998. Is price dispersion the result of careful planning or competitive forces? *Review of Industrial Organization,*13, 523-541.

Ito, H., & Lee, D. (2007). Domestic codesharing, alliances and airfares in the US airline industry. *Forthcoming in Journal of Law and Economics.*

Williams, K. R. (2013). *Dynamic airline pricing and seat availability.* Department of Economics, University of Minnesota (Working Paper).

KIMES, S. E. (1997). *Yield Management: An Overview.* Yield Management, Strategies for the Service Industries. London: Routledge.

Kimes, S. E. (1989). Yield management: a tool for capacity-constrained service firm. *Journal ofOperations Management* 8, 348-363.

Lee, T. C., & Hersh, M. (1993). A model for dynamic airline seat inventory control with multipleseat bookings. *Transportation Science,* 27, 252-265.

Littlewood, K. (1972). Forecasting and control of passengers. *12th AGIFORS Symposium Proceedings. Nathanya, Israel,* (pp.95-128).

Fedorco, L., & Hospodka, J. (2013). *Airline Pricing Strategies in European Airline Market*, 7(2).

McGill, J., & van Ryzin, G. (1999). Revenue management: Research Overview and Prospects. *Transportation Science,* 33, 233-256.

Netessine, S., & Shumsky, R. (2004). Revenue management games: Horizontal and vertical competition. *Management Science,* 51(5) 813-831.

Peteraf, M. A., & Reed, R. (1994). Pricing and performance in monopoly airline markets. *Journal of Law and Economics*, 37(1), 193-213.

Robinson, L. W. (1995). Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes. *Operations Research,* 43, 252-263.

Rothman, A. (2006) Sep. 1. Air France-KLM raises profit forecast for 2006 MARKETPLACE by Bloomberg. *International Herald Tribune,* page 13.

Shumsky, R. (2006). The southwest effect, airline alliances, and revenue management. *Journal of Revenue and Pricing Management*, 5(1), 83-89.

Talluri, K., & x van Ryzin, M. (2004). *The theory and practice of revenue management.* Kluwer Academic Puslishers.

Vulcano, G., van Ryzin, G., & Chaar, W. (2010). Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, 12 (3), 371-392.

Windle, R., & Dresner, M. (1993). Competition at “Duopoly” Airline Hubs in the US.

*Transportation Journal,*33(2), 22-30.

Wollmer, R. D. (1989). An airline seat management model for a single flight leg route when lower fare classes book first. *Operations Research,* 40, 26-37.

Wright, C. P., Groenevelt, H., & Shumsky, R. (2006). Dynamic Revenue Management in Airline

Alliances, working paper.

Zhao, W., & Zheng, Y. (2000). Optimal dynamic pricing for perishable assets with non homogeneous demand. *Management Science,* 46, 375-388.

Appendix 1

ANOVA test results (main effect) of Independent variables on airline ticket price between two destinations.

S.No.	Factor	Description	F Value	P Level
1	PAX	Number of Passengers	9.289	0.0105
2	Dep Station	Departure Station	6.432	0.0400
3	Arr Station	Arrival Station	1.560	0.0827(NS)
4	Days to Departure	Difference between Booking Date and Departure Date	7.524	0.0295
5	Booking Type	Booking Channel	5.555	0.0474
6	Fare Basis	Fare Basis	3.786	0.0505
7	Class of Service	Code for Class of the ticket	6.885	0.0360
8	ASK	Available Seat Kilometer	0.500	0.2466(NS)
9	Comp_Scenario	Number of competitors Operating between the same stations	9.694	0.0077
10	RPK	Revenue Passenger Kilometer	6.831	0.0360
11	No. of Flights	Number of available flights on the same day	8.467	0.0248
12	Promo Code	Code for Promotions in the corresponding Channel	3.544	0.0621
13	Base Price	Base fare	7.303	0.0343
14	Convenience Fee	Convenience Fee	1.694	0.0790(NS)
15	UDF	User Development Fee	8.972	0.0222
16	AAT	Airport Authority Tax	7.898	0.0283
17	PSF	Passenger Service Fee	0.552	0.1708(NS)
18	GST	Government Service Tax	0.148	0.3734(NS)
19	CUTE Fees	Transaction Fees	1.103	0.0892(NS)
20	Fuel Charge	Fuel Charges	0.684	0.0898(NS)
21	Capacity	Total available Capacity	8.120	0.0248
22	Currency Code	Currency Code	7.495	0.0299
23	STD	Scheduled Time of Departure	5.586	0.0461
24	STA	Scheduled Time of Arrival	9.511	0.0104
25	Flight Tail No.	Registration Number of the Flight	3.876	0.0504
26	L.F	Load factor	2.423	0.0754

NS – Not Significant

Forecasting The Stock Market Values Using Hidden Markov Model

R. Sasikumar*, A. Sheik Abdullah*

Abstract

The financial market influences personal corporate financial lives and the economic health of a country. Price change of stock market is not a completely random model. The pattern of financial market has been observed by some economists, statisticians and computer scientists. This paper gives a detailed idea about the sequence and state prediction of stock market using Hidden Markov Model and also making inferences regarding stock market trend. The one day difference in close value of stock market value has been used for some period and the corresponding transition probability matrix and emission probability matrix are obtained. Seven optimal hidden states and three sequences are generated using MATLAB and then compared.

Keywords: Hidden Markov Model, Transition Probability Matrix, Emission Probability Matrix, Stock Market, States and Sequence

Introduction

The most of the trading in Indian stock market is classified in two categories, the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The BSE has been functioning since 1875. The NSE was founded in 1992 and started trading in 1994. Even though both exchanges have the same trading mechanism, trading hours, settlement process, etc., they are having high demand from people. The two prominent Indian market indices are Sensex and S&P CNX Nifty.

Financial market (Stock Market) is a platform for investors to own some shares of a company. Investors will then become a part of the company members and share in both gains and losses of that particular company. This is a better way for the investors to get extra income apart from their regular salary. Changes of share prices on every day make it more volatile and difficult to predict the future price. When purchasing a stock, it does not guarantees to give anything in return. Thus, it makes stocks risky in investment, but investors can also get high profit return. When investors take wrong decision in choosing the counters, it may end up in capital loss. The behavior of stock market returns has been deeply discussed over some years. In this paper, the hidden states and sequence are generated for stock market values using Hidden Markov Model (HMM) through software.

Review of related works

There are so many researches going on stock market analysis. Rabiner (1989) used precise HMMs, in which the state sequence estimation problem can be solved very efficiently by the Viterbi algorithm whose complexity is linear in the number of nodes, and quadratic in the number of states. However, this algorithm only emits a single optimal (most probable) state sequence, even in cases where there are multiple (equally probable) optimal solutions. Hassan and Baikunth Nath (2005) used HMM to predict next day closing price for some of the airlines. They considered four input attributes for a stock, and they were the opening price, highest price, lowest price and closing price. These four attributes of previous day were used to predict next day's closing price. Hassan (2009) introduced the new combination of HMM and Fuzzy model to forecast the stock market data. He classified the data set as daily opening, high, low and closing prices to predict the next day's closing price. HMM-fuzzy model is more reliable and profitable than the other model.

Jyoti Badge (2012) used Macro-Economic factor as a technical indicator, which is used to identify the patterns of the market at a particular time. For selecting technical indicator author was applying principal component

* Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India.
Email: sasikumarmsu@gmail.com and sheik.stat@gmail.com

analysis. Luca et al., (2013) investigated the dynamic patterns of stock markets by exploiting the potential of the HMM for defining different market regimes and providing transition probabilities for regime-switching. Tuyen (2013) used HMM to estimate the parameter of the Markov Black-Sholes model to predict the option prices in the stock market. The historical data of daily VN-Index (Vietnam Stock Market) were taken from 2009 to 2011 for finding the four hidden states corresponding to the Normal distribution $N(\mu_i, \sigma_i)$ for $i = 1,2,3,4$ with the help of HMM. Kai Cui (2014) explained that, the variation of financial time sequence for Shangai composite index was predicted by introduction of a dual state HMM. He also justified that the HMM was the best tool to predict the variation of financial time sequence. Somani et al., (2014) surveyed support vector machine, neural network and HMM in the area of stock market forecasting. HMM is more efficient in getting information from the result, showing future behavior of stock market valuesand fluctuations.

Methodology

Hidden Markov Model

HMM has been successful in analyzing and predicting phenomena's relying on a time dependence or time series. It is very effective and intuitive approach to many sequential pattern recognition tasks, such as speech recognition, protein sequence analysis, machine translation, pair wise and multiple sequence alignments, gene annotation, classification and similarity search.

A HMM is a doubly stochastic process in which an underlying stochastic process is unobservable, which means that the state is hidden. This can only be observed through another stochastic process that produces a sequence of observations. Thus, if $S = \{S_n, n=1, 2,...\}$ is a Markov process and $F = \{F_k, k=1, 2, \ldots\}$ is a function of S, then S is a hidden Markov process or HMM that is observed through F, and S is also known as the state process that is hidden and F as the observation process that can be observed. The observed event is called as a "symbol" and the invisible factor underlying the observation a "state".

A HMM is usually defined as a 5-tuple (S, F, P, ψ, π), where

$S = \{s_1, s_2, \ldots, s_n\}$ is a finite set of n states.

$F = \{o_1, o_2, \ldots, o_m\}$ is a finite set of m possible symbols.

$P = \{p_{ij}\}$ is the set of state-transition probabilities, where p_{ij} is the probability that the system goes from state s_i to state s_j.

$\psi = \{\psi_i(o_k)\}$ are the observation probabilities, where $\psi_i(o_k)$ is the probability that the symbol o_k is emitted when the system is in state s_i.

$\pi = \{\pi_i\}$ are the initial state probabilities; that is the probability that the system starts in state s_i.

As the states and the output sequence are understood, it is usually denoted by the parameters of a HMM by $\lambda = (P, \psi, \pi)$.

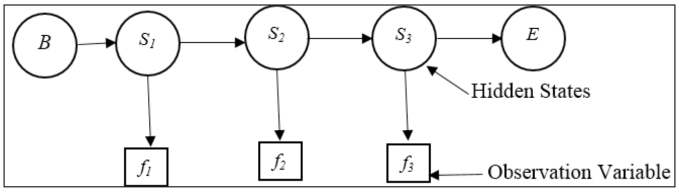


Figure 1: General Structure of a Hidden Markov Model

From the Figure 1, the $S_i$are the hidden states that is to be estimated and the F_i are the observation of the random variables from which the S_i are to be estimated. The letters B and E indicate the beginning and end of the sequence of states.

Transition Probability Matrix

The transition probability P_{jk},where $P_{jk} \geq 0$, for all j. These probabilities may be written in the matrix form,

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

This is called the transition probability matrix (tpm). P is a stochastic matrix ie.a square matrix with non-negative elements and row total is equal to one.

Materials and Methods

In this paper, Oil India Ltd is sample share, its daily close value data for two months period is considered. Three observing symbols I, N and D are indicated. The symbol I-indicates Increasing, N-indicates No change and D-indicates Decreasing. If n^{th} day's close value – (n-1) th day's close value > 0, then observing symbol is I. If n^{th} day's close value – $(n-1)^{th}$ day's close value < 0, then observing symbol is D. If n^{th} day's close value – $(n-1)^{th}$ day's close value = 0, then observing symbol is N.

Seven hidden states are assumed and are denoted by the following symbols S_1, S_2, S_3, S_4, S_5, S_6, S_7

where,

S_1 - very low

S_2 - low

S_3 - moderate low

S_4 - no change

S_5 - moderate high

S_6 – high

S_7 – very high

Since the above mentioned states are not directly observable, in this situation the stock market values are considered as hidden. From the hidden state sequences, it is possible to produce the observations.

The various probability values of tpm and emission probability matrix (epm) for difference in one day, two days and three days close values are calculated as follows: tpm and epm for one day close value difference

$$
\begin{array}{c|ccccccc}
 & S1 & S2 & S3 & S4 & S5 & S6 & S7 \\
\hline
S1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
S2 & 1/5 & 2/5 & 1/5 & 0 & 0 & 0 & 1/5 \\
S3 & 0 & 0 & 0 & 1/11 & 8/11 & 2/11 & 0 \\
S4 & 0 & 1/4 & 2/4 & 1/4 & 0 & 0 & 0 \\
S5 & 0 & 1/14 & 7/14 & 1/14 & 3/14 & 2/14 & 0 \\
S6 & 0 & 0 & 1/6 & 0 & 4/6 & 0 & 1/6 \\
S7 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0
\end{array}
$$

Figure 1(a): tpm

Table 1: The Closing Value of a Stock Market Tpm and Epm for two Day Close Value Difference

S. No	Close Value	Diff in 1 day	Observation Symbol	Diff in 2 day	Observation Symbol	Diff in 3 day	Observation Symbol
1	458.00						
2	465.70	7.70	I				
3	469.00	3.30	I	-4.40	D		
4	467.35	-1.65	D	-4.95	D	-0.55	D
5	469.90	2.55	I	4.20	I	9.15	I
6	473.65	3.75	I	1.20	I	-3.00	D
7	466.00	-7.65	D	-11.40	D	-12.60	D
8	487.20	21.20	I	28.85	D	40.25	I
9	501.00	13.80	I	-7.40	D	-36.25	D
10	518.85	17.85	I	4.05	I	11.45	I
11	508.60	-10.25	D	-28.10	D	-32.15	D
12	500.15	-8.45	D	1.80	I	29.90	I
13	478.45	-21.70	D	-13.25	D	-15.05	D
14	478.45	0.00	N	21.70	I	34.95	I
15	467.50	-10.95	D	-10.95	D	-32.65	D
16	456.30	-11.20	D	-0.25	D	10.70	I
17	451.95	-4.35	D	6.85	I	7.10	I
18	457.05	5.10	I	9.45	I	2.60	I
19	459.25	2.20	I	-2.90	D	-12.35	D
20	458.50	-0.75	D	-2.95	D	-0.05	D
21	458.50	0.00	N	0.75	I	3.70	I
22	458.50	0.00	N	0.00	N	-0.75	D
23	453.65	-4.85	D	-4.85	D	-4.85	D
24	457.00	3.35	I	8.20	I	13.05	I

S. No	Close Value	Diff in 1 day	Observation Symbol	Diff in 2 day	Observation Symbol	Diff in 3 day	Observation Symbol
25	456.75	-0.25	D	-3.60	D	-11.80	D
26	467.65	10.90	I	11.15	I	14.75	I
27	470.10	2.45	I	-8.45	D	-19.60	D
28	470.95	0.85	I	-1.60	D	6.85	I
29	483.30	12.35	I	11.50	I	13.10	I
30	481.30	-2.00	D	-14.35	D	-25.85	D
31	483.45	2.15	I	4.15	I	18.50	I
32	478.65	-4.80	D	-6.95	D	-11.10	D
33	480.65	2.00	I	6.80	I	13.75	I
34	477.50	-3.15	D	-5.15	D	-11.95	D
35	485.95	8.45	I	11.60	I	16.75	I
36	487.55	1.60	I	-6.85	D	-18.45	D
37	486.00	-1.55	D	-3.15	D	3.70	I
38	492.60	6.60	I	8.15	I	11.30	I
39	491.15	-1.45	D	-8.05	D	-16.20	D
40	493.70	2.55	I	4.00	I	12.05	I
41	493.70	0.00	N	-2.55	D	-6.55	D
42	488.35	-5.35	D	-5.35	D	-2.80	D
43	490.00	1.65	I	7.00	I	12.35	I
44	499.00	9.00	I	7.35	I	0.35	I
45	501.25	2.25	I	-6.75	I	-14.10	D

$$
\begin{array}{c}
\\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7
\end{array}
\begin{array}{c}
\begin{array}{ccc} I & N & D \end{array} \\
\begin{bmatrix}
0 & 1 & 0 \\
1/5 & 0 & 4/5 \\
10/11 & 1/11 & 0 \\
0 & 1/4 & 3/4 \\
5/14 & 1/14 & 8/14 \\
1/6 & 0 & 5/6 \\
0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 1(b): **epm**

$$
\begin{array}{c}
\\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7
\end{array}
\begin{array}{c}
\begin{array}{cc} I & D \end{array} \\
\begin{bmatrix}
1 & 0 \\
1 & 0 \\
14/17 & 3/17 \\
0 & 1 \\
2/14 & 12/14 \\
0 & 1 \\
0 & 1
\end{bmatrix}
\end{array}
$$

Figure 2(b): **epm**

$$
\begin{array}{c}
\\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & S6 & S7 \end{array} \\
\begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1/4 & 0 & 1/4 & 0 & 2/4 \\
0 & 0 & 5/17 & 0 & 9/17 & 3/17 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1/14 & 2/14 & 7/14 & 1/14 & 3/14 & 0 & 0 \\
0 & 1/3 & 2/3 & 0 & 0 & 0 & 0 \\
0 & 1/2 & 1/2 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

Figure 2(a): **tpm**

tpm and epm for three day close value difference

$$
\begin{array}{c}
\\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & S6 & S7 \end{array} \\
\begin{bmatrix}
0 & 0 & 0 & 0 & 2/4 & 1/4 & 1/4 \\
0 & 0 & 1/9 & 0 & 5/9 & 1/9 & 2/9 \\
0 & 0 & 3/7 & 0 & 4/7 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
2/16 & 6/16 & 3/16 & 0 & 5/16 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
2/3 & 1/3 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

Figure 3(a): **tpm**

$$
\begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \end{array}
\begin{array}{c} \begin{array}{cc} \mathrm{I} & \mathrm{D} \end{array} \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 5/7 & 2/7 \\ 1 & 0 \\ 5/16 & 11/16 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \end{array}
$$

Figure 3(b): epm

From the above TPM and EPM hidden states and sequence have been generated using MATLAB software. Difference of one day, two day and three day hidden states and sequence are given below respectively. From the sequence and states we can predict the future values of stock value.

1	Sequence:	D	I	D	I	I	D	D	I	D	I
	States:	S4	S2	S7	S6	S5	S5	S5	S3	S5	S3
2	Sequence:	D	I	D	I	D	I	D	I	I	I
	States:	S5	S3	S5	S3	S6	S3	S5	S2	S3	S3
3	Sequence:	I	I	I	D	D	I	D	I	I	I
	States:	S5	S5	S3	S5	S5	S2	S6	S2	S5	S3

Conclusion

Stock market values are unpredictable because of the variation of several factors. So there is no single method which can perfectly forecast the stock price values, HMM is no exception. Even though through this paper, the HMM model easily recognized three states of the stock market and also it was used to forecast the future values. In this paper, hidden states and sequences have been generated to identify, so that, we can easily identify the future states and also easily identify the sequence whether the next day value is increasing or decreasing and increasing/decreasing level can also be observed. We can identify whether the increasing level is moderate, high or high or very high and also decreasing level whether moderate or low or very low. This is very useful for short term as well as long term investors.

Acknowledgement

The authors acknowledge University Grants Commission for providing the infrastructure facility under the scheme of SAP (DRS-I). The second author acknowledges UGC for financial support to carry out this research work under Basic Science Research Fellowship.

References

Hassan, M. R., & Nath, B. (2005). Stock Market Forecasting Using Hidden Markov Model: A New Approach, *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, 192-196.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications*, 33(1), 171-180.

Hassan, M. R. (2009). A combination of hidden Markov model and fuzzy model for stock market forecasting. *Neurocomputing*, 72, 3439-3446.

Badge, J. (2012). Forecasting of Indian Stock Market by effective macro-economic factors and stochastic model. *Journal of Statistical and Econometric Methods*, 1(2), 39-51.

Cui, K. (2014). Analysis of Financial Time Sequence using Hidden Markov Models. *Journal of Multimedia*, 9(6), 810-815.

Angelis, L. D., & Paas, L. J. (2013). A dynamic analysis of stock markets using a hidden Markov model, *Journal of Applied Statistics*, 40(8), 1682-1700.

Ibe, O. C. (2009). *Markov processes for stochastic modeling.* Elsevier Academic Press

Rabiner, L. R., & Juang, B. H. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine January*, 4-16.

Rabiner, L. R. (1989). A tutorial on HMM and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.

Somani, P., Talele, S., & Sawant. S. (2014). Stock market prediction using Hidden Markov model. *IEEE*, 89-92.

Tuyen, L. T. (2013). Markov financial model using hidden Markov model. *International Journal of Applied Mathematics and Statistics*, 40(10), 72-83.

Weigend, S., & Back, A. D. (1998). What drives stock returns? An independent component analysis. *In Proceedings of the IEEE/IAFE/INFORMS 1998 Conference on Computational Intelligence for Financial Engineering, IEEE*, New York., pp. 141- 156.

Zhang, D., & Xiaomin, Z. (2009). Study on forecasting the stock market trend based on stochastic analysis method. *International Journal of Business and Management*, 4(6), 163-170.