Forecasting Hospital Emergency Department Arrivals

Bernard J. Morzuch
Department of Resource Economics, University of Massachusetts, Amherst, MA 01003 USA
morzuch@resecon.umass.edu

P. Geoffrey Allen
Department of Resource Economics, University of Massachusetts, Amherst, MA 01003 USA
allen@resecon.umass.edu

Abstract

Predicting demand for hospital emergency department services is contingent upon the arrival distribution of incoming patients. We focus on one particular hospital's emergency department and its arrival distribution on an hourly basis. Accurate hourly forecasts are important guides for physician and nurse staffing, both of which are major components in health care capacity planning.

In a previous study, we forecasted hourly arrivals for each day using standard Holt-Winters (H-W) exponential smoothing methods. We picked up movements in actual arrivals for 168 hours ahead; however, we under-predicted peaks and troughs. Poor predictions were the result of not modeling all relevant levels of seasonality. We modeled daily effects but ignored weekly effects.

In the present study, we compare forecasts from standard H-W having daily seasonality with: (1) a double seasonal exponential smoothing method that allows the inclusion of one cycle nested within another; and (2) an unobserved components model.

We reserve 3360 hourly observations at the end of the series (which amounts to 20 weeks of hourly data) to evaluate forecasting performance. We evaluate forecasting performance by (1) performing a series of one-through-168 steps-ahead forecasts at the end of the initial within-sample period; (2) updating the within-sample period by one observation (but doing no re-estimation) and again performing a series of one- through-168 steps-ahead forecasts; (3) repeating the updating process until running out of post-sample observations; (4) gathering all one-step ahead forecasts and calculating RMSE; (5) repeating the RMSE calculation for each category of steps-ahead forecasts; (5) applying this procedure to each forecasting approach; and (6) comparing RMSEs over the different forecasting horizons for each approach.

By comparing RMSE for each forecast horizon and for each method, we can judge the forecasting performance of these more flexible forecasting approaches relative to standard H-W.

Presented at the 26th Annual Symposium on Forecasting
Santander, Spain, June 11-14, 2006

Forecasting Hospital Emergency Department Arrivals

Introduction

One area of interest in hospital Emergency Department (ED) research is the "refined ability to forecast and model ED and hospital unscheduled services" Mahapatra, *et al.*, (2004). Surprisingly, modeling in this area using time-series methodologies has been limited or, more appropriately, nonexistent.

Predicting resource intensity, i.e., the timely application of needed physician and nursing expertise, is dependent on accurately forecasting patient arrival patterns. An ED is typically a 24-hours-a-day/seven-days-a-week/365-days-per-year enterprise. The ideal is to predict arrivals on an hourly basis for each of the 168 hours during the week. Accurate arrival predictions provide information relating to the medical resources needed to accommodate patient disposition.

In a previous study, we forecasted hourly arrivals for each day using standard Holt-Winters (H-W) exponential smoothing methods. We picked up movements in actual arrivals for 168 hours ahead; however, we under-predicted peaks and troughs. Poor predictions were the result of not modeling all relevant levels of seasonality. We modeled intra-day patterns but ignored intra-week patterns. The purpose of this study is to correct this deficiency.

Emergency Department Arrivals: York Hospital

One does not need to make an appointment for services at an ED. For any day of the week, arrivals can occur at any hour. The ideal is to discover if there is a pattern in arrivals during a day and then to determine whether or not such patterns, if they do exist, are different among the seven days.

To explore this issue, we were fortunate to have hourly arrivals to the Emergency Department at York Hospital in York, Pennsylvania. Specifically, we had 160,471 arrivals during the period November 11, 1999 to June 28, 2002. Along with other detailed information, each individual's arrival time was presented in the original data file. Prior to using the data for analysis, we placed each arrival into a time bin and assigned it a corresponding time-bin integer. For example, if an individual arrived between 12:00 a.m. and 12:59 a.m. (which we call time bin 1), this particular arrival is assigned the integer 1. If an individual arrived between 11:00 p.m. and 11:59 p.m., (which we call time bin 24), this particular arrival is assigned the integer 24.

Descriptive Account Of Arrivals

Figure 1 presents a plot of the *total* number of arrivals by hour of the day over the entire period. No attempt is made to distinguish among the different days of the week in this figure. Here, we simply are trying to visualize which hours of the day result in the most arrivals and which result in the least. In general, it appears that most traffic is between hours 11 and 22, i.e., between 10:00 a.m. and 9:59 p.m. Traffic is relatively light (but is taking an upturn) before 10:00 a.m. and relatively light again (and is taking a downturn) after 9:59 p.m.

Figure 2 shows the pattern in *total* hourly arrivals for each day of the week over the entire period. This figure simply disaggregates the information in Figure 1 by day of the week. Notice that the pattern presented in Figure 1 holds roughly for the individual days of the week in Figure 2. Figure 2 also shows that Sunday, Saturday, and Monday account for the highest levels of arrivals.

While Figure 2 shows the pattern in total arrivals per hour for each day over the entire period, Figure 3 provides actual numbers relating to total daily arrivals over the entire period. Specifically, Figure 3 shows that Sunday leads the pack with 25,416 total arrivals followed by Saturday with 24,351 total arrivals and Monday with 23,374 total arrivals. Each of the remaining four days hovers around 22,000 total arrivals. Please notice that each total is calculated over the entire period. Also, weekends are obviously associated with high ED activity.

Next, we look at each day-of-the-week's total arrivals from a different perspective. Consider that there are approximately 137 Sundays in our data set. Each Sunday generates a total number of arrivals. If we sum all Sunday total arrivals and divide by 137, we get the average total arrivals for Sunday. This figure – 185.52 – is reported in the upper left corner of Table 1. The full catalog of descriptive statistics of each day's total arrivals is presented in Table 1. The last column provides information over the entire period.

Time-Series Characteristics Of Hourly Arrivals

We begin analyzing the series by asking three basic questions relating to its general characteristics: (1) Does there appear to be some sort of *seasonal pattern,* and, if so, how is this seasonal pattern defined? (2) Over the duration of the series, does there appear to be some sort of *trend*; i.e., either stochastic or deterministic? (3) Does the series appear to revert to a particular *level*? Honing in on our data set's characteristics as posed by these three questions will assist with the selection of a modeling procedure.

Figure 2 provides a preliminary answer to Question (1). First, there appears to be a pattern in a given day's 24-hour series. Secondly, this pattern is consistent from one day to the next. Linking all days' 24-hour series would suggest some sort of daily model with an hourly "seasonal" pattern; i.e., a model characterized by hourly seasonality.

We can get an idea of the answer to Question (2) by plotting the series over its duration; i.e., from November 11, 1999 to June 28, 2002. We have two choices regarding the plot. We can plot 24 separate daily arrival observations for each day of the year over the entire period, which will result in a graph that looks like a solid and thick horizontal bar (due to the vast number of observations). In turn, the plot would be so busy that discerning a visual pattern would be impossible. As an alternative, we can plot the daily total number of arrivals over the entire period, which will still be busy-looking but nowhere to the degree as with the first alternative. This is presented as Figure 4.

Figure 4 is very telling in that it provides an indication that daily arrivals are trending neither upward nor downward over the duration of the data set; i.e., there appears to be no trend. Alternatively, the overall horizontal appearance of the plot likewise suggests that arrivals are relatively level over the duration of the data set.

The height of individual spikes in Figure 4 is indicative of daily variability. There are two reasons for this variability: (1) certain days of the week have more daily arrivals than other days; and (2) special "event" days, e.g., holidays, may lead to abnormally high ED arrivals.

Time Series Tests

As a next step and prior to model development, we performed several standard tests on and calculated several important statistics for the arrivals series. We tested and rejected the null hypothesis that the series itself is white noise. This provided evidence that the series should be modeled. As a way of addressing the possibility of nonstationarity of the series due to trend, we performed a unit root test. We rejected the null hypothesis of nonstationarity due to trend.

We plotted the autocorrelations and partial autocorrelations of the arrivals series. The correlogram itself (not shown here) looks like a classic sine wave. The wavelength (measured from one crest to the next) is 24 periods or hours. Even beyond 48 hours, the amplitude (which measures the height or strength) maintains a constant level; it does not damp. The obvious conclusion is that the series has a strong seasonal pattern that does not change.

Within-Sample Versus Post-Sample Data Sets

Prior to model development we divided the data set into two parts. The first part comprised that portion of the data set that would be used in model development. This covered the period 14 November 1999 at 12:00 a.m. through 8 February 2002 at 11:59 p.m. We refer to these observations as the within-sample data set or the observations used for model construction.

The second part of the data set consisted of the observations from 9 February 2002 at 12:00 a.m. through 28 June 2002 at 11:59 p.m. This amounted to 3360 hourly observations or 20 weeks of hourly data. These observations were not used in the development of the model. We refer to these observations as the holdout sample or post-sample data set.

The holdout sample serves a useful purpose. It gauges how well the model performs on a “fresh” data set. Notice that the model did not use these observations in its development. Thus, each actual observation for any one of the 3360 hours in the holdout sample can be compared against the matching hourly prediction obtained from using the model. This permits us to assess how well the model predicts.

Model Development

In a previous study, we considered three broad categories of models: (1) regression-based; (2) autoregressive integrated moving average (ARIMA); and (3) exponential smoothing. The behavior of the seasonal, trend, and level components of the arrivals series provided us with guidance relating to model specification in each broad category. In that study, we found that standard Holt-Winters exhibiting daily seasonality performed better than other exponential smoothing methods and better than models in the other two broad categories. Again, our guidance in this choice was RMSE in the post-sample period.

Amended Modeling Procedures

A plot of our hourly data reveals two seasonal cycles. First, there is the within-day (intra-day) 24-hour seasonal cycle; i.e., this type of seasonality completes itself during each 24-hour period. Secondly, there is the within-week (intra-week) 168-hour seasonal cycle; i.e., this type of seasonality completes itself during each week. Effectively, the daily cycle is nested within the weekly cycle.

Time-series models that attempt to capture both types of seasonality are commonly referred to as double-seasonal. We consider two approaches for modeling our daily and weekly cycles: the Unobserved Components Model (UCM) approach and double-seasonal exponential smoothing.

Unobserved Components Model

UCM is a special case of time series modeling in state space form. State space form consists of a measurement equation and a transition equation. Within each equation is a properly defined state vector and one or more system matrices. In the UCM framework, typical components of the state vector in the measurement equation include trend, seasonal, cycle, and irregular components. Each component of the time series is separately modeled as stochastic and, thus, unobserved. Hence, we have the name Unobserved Components Model.

. The general UCM representation in additive form is:

$$y_t = \mu_t + \gamma_t + \psi_t + r_t + \sum_{j=1} \beta_j x_{t-j} + \sum_{i=1} \phi_i y_{t-i} + \varepsilon_t \tag{1}$$

where μ_t is a combination of level and trend components, γ_t is the seasonal component, ψ_t is the cyclical component, and r_t is an autoregressive component. Regression terms are represented by $\sum_{j=1} \beta_j x_{t-j}$, and lagged regression terms are represented by $\sum_{i=1} \phi_i y_{t-i}$.

Finally, ε_t is the irregular component or error term.

We can take advantage of the UCM approach to model multiple seasonal patterns because of the flexible and structural nature of the state space form. The UCM representation for our particular case simplifies compared to the specification in Equation (1) above. Consider that there are 168 hours in a week. Thus, an hour is our time unit, and time t varies from 1 to 168. The components of our model are: (a) μ_t, a level term *without* trend during hour *t*; (b) $\gamma_{i,t}$, a seasonal component identified uniquely with hour *i* in each 24-hour period, with *i* = 1,2, …, 24; and (c) λ_{jt}, a seasonal component identified uniquely with day *j* in each week, with *j* = 1,2, …, 7.

Given this information, our model becomes:

$$y_t = \mu_t + \gamma_{1,t} + \gamma_{2,t} + ... + \gamma_{24,t} + \lambda_{1,t} + \lambda_{2,t} ... + \lambda_{7,t} + \varepsilon_t \qquad (2)$$

The recursion formula for level is:

$$\mu_t = \mu_{t-1} + \xi_t \qquad (3)$$

The recursion equations for the remaining components are presented in Harvey (pp. 104-109).

Double Exponential Smoothing

Taylor, *et al*., (2006) forecast hourly demand for electricity using double-seasonal ARIMA and double-seasonal exponential smoothing, among other approaches. The appeal of double-seasonal exponential smoothing is that it involves no model specification. The method is an extension of standard Holt-Winters exponential smoothing; it accommodates an additional seasonal cycle. This requires an extra smoothing equation and an additional seasonal index. Typical smoothing equations for double multiplicative seasonality are as follows (See Taylor, *et al*., p. 6.):

$$L_t = \alpha(y_t / (D_{t-s_1} W_{t-s_2})) + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad (4)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \qquad (5)$$

$$D_t = \delta(y_t / (L_t W_{t-s_2})) + (1-\delta)D_{t-s_1} \qquad (6)$$

$$W_t = \omega(y_t / (L_t D_{t-s_1})) + (1-\omega)W_{t-s_2} \qquad (7)$$

where L_t and T_t are smoothed level and trend, respectively; D_{t-s_1} and W_{t-s_2} are the seasonal indices for the intra-day and intra-week seasonal cycles, respectively. Notice that $s_1 = 24$ and $s_2 = 168$; α, β, δ, and ω are smoothing parameters.

The h-step-ahead forecasting equation is:

$$\hat{y}_t(h) = (L_t + hT_t)D_{t-s_1+h}W_{t-s_2+h} \qquad (8)$$

Model Selection And Performance

Within-sample estimation for each approach covered the period 14 November 1999 at 12:00 a.m. through 8 February 2002 at 11:59 p.m. Parameter estimates and other estimation results are not presented.

We evaluate forecasting performance using the following strategy: (1) at the end of the initial within-sample period and for each method, we forecast hourly arrivals for one-through-168 steps ahead, being cognizant that we have 3360 reserved observations in the post-sample data set; (2) after generating the forecasts above, we then update the initial within-sample data set with the first post-sample observation. This also means that the post-sample moves forward by one observation. Thus, after the initial forecasting exercise, we have a revised starting value for the first post-sample observation to be used in a new round of forecasting; (3) as within-sample and post-sample data sets are updated, we do not re-estimate the smoothing parameters; i.e., we update without re-estimating. With each update, we have a new recursion level that is used throughout the new one-through-168 steps-ahead forecasts. Seasonal factors likewise are not re-estimated; they are historical and appropriate to the specific time period; (4) with the ground rules presented as to how the model is to be used to obtain the forecasts, we repeat the first step; i.e., we forecast hourly arrivals for one-through-168 steps ahead based upon moving the post-sample one period forward; (5) we keep moving the post-sample forward in increments of one and generating forecasts until we run out of observations; (6) when the process is complete, we collect the entire series of one-step-ahead forecasts and calculate RMSE for all one-step-ahead forecasts; we do the same for all two-steps-ahead forecasts, for all three-steps-ahead forecasts, and so on, until we complete the process for all 168-steps-ahead forecasts; (7) as indicated at the outset, we apply this process to each forecasting approach; and, finally, (8) we compare RMSEs over the different forecasting horizons for each approach.

Results

As indicated in the previous section, we focused on RMSE calculations based on one-to-168 steps-ahead forecasts. Accomplishing this task is effectively a programming challenge. Understanding the behavior of the components that drive the calculations makes this exercise feasible.

We faced one difficulty. While we were able to complete this task for standard H-W and double seasonal exponential smoothing, we were able to generate only one-step-ahead forecasts for the Unobserved Components Model. We had difficulty interpreting output for UCM components that was needed to do forecasts for anything beyond one step ahead. Solving this dilemma is left as a future exercise.

Figure 5 provides comparisons between the RMSEs for H-W and double seasonal exponential smoothing. The first item to notice is that double seasonal RMSEs are consistently below H-W RMSEs over all forecast horizons. This is consistent with our expectations; i.e., since arrivals have an hourly pattern and a day-of-the-week pattern, a method that accounts for both ought to perform better.

Figure 5 also shows that H-W RMSEs rise as the forecast horizon increases, as would be expected. But then, quite curiously, at a horizon of approximately 68 they begin to fall, and the drop continues through horizon 124. After that, they sort of bounce around in a horizontal fashion. We would not have expected the RMSEs to fall as the forecast horizon increased.

Although the graph gives the appearance of a dramatic increase followed by a dramatic drop, we should point out that our minimum RMSE is approximately 2.410, and our maximum RMSE is approximately 2.484. Thus, the RMSE range over all forecast horizons is approximately 0.074. (This amounts to three percent of the minimum RMSE value). Nevertheless, it is still curious that RMSEs decline.

We are not accustomed to seeing studies that make comparisons for so many (i.e., 168) forecast horizons. Taylor, *et al*, (2006, p.10), for example, have a horizon of 24 using MAPE for their electricity data. As their horizon increases, MAPE likewise increases and eventually levels off around horizon 21 and stays level through horizon 24. Returning to Figure 5, if we were to restrict ourselves to only 24 forecast horizons and if we were able to use MAPE as a performance measure, it is conceivable that our graphs would have an appearance that resembles Taylor, *et al.* (2006) more closely.

Finally, recall that we stated that we could analyze the performance of UCM only for one-step-ahead forecasts and for no forecast horizons beyond one period. With UCM, we found that RMSE was 2.516 based on a forecast horizon of one period. With H-W, RMSE was 2.413, and with double seasonal, RMSE was 2.403. This is indeed a curious result. At the minimum, we would have expected UCM to outperform standard H-W, particularly because of the reasons previously given.

James Taylor made available to us his algorithm that performs double seasonal exponential smoothing. We amended it quite straightforwardly to do standard H-W. Taylor's forecasting equation (See Taylor, *et al*., 2006, p. 6.) has a parameter to adjust for first-order autocorrelation. We mention this because we did not accommodate this possibility in our UCM. Although we found no evidence of autocorrelation with our arrivals data as did Taylor, *et al*. (2006) with their electricity data, we could have estimated a UCM with an autocorrelation component to be consistent with the algorithm used for H-W and double seasonal.

Discussion

It is apparent from this exercise that accommodating additional layers of seasonality improves forecast performance when these layers do indeed exist. Accessing a data set with characteristics like this was new to us. This newness forced us to focus on the interpretation that is to be given to the parameters that are estimated from the procedure.

We paid attention to issues raised by other researchers who used data sets with similar characteristics in order to gain additional insight. Taylor, *et al*. (2006) experienced a strong autocorrelation pattern with their electricity data. For their two sets of data, their AR parameter estimates were 0.88 and 0.98, respectively. We experienced no such pattern with our arrivals data. Our AR parameter estimate was 0.093. In fact, the level, trend, daily seasonal, and weekly seasonal indices were dramatically different in size between the two studies.

When the AR parameter estimate is close to 1, as with the data used by Taylor, *et al.* (2006), the last error is effectively added to each of the 1 through h-steps-ahead forecasts. There is little damping effect through time. The effect is almost that of an intercept correction. This situation is in contrast to our ED arrivals data. Our data are quite stable over time. Our AR parameter estimate is much smaller, and its effect disappears almost immediately.

Clearly the focus of this study is accommodating the possibility of double seasonality. For ED arrivals, this leads to forecasting improvement. Equally important is the size of the smoothing parameter estimates. Size of parameter estimates has important implications regarding how past information is carried forward to future forecast horizons.

References

Harvey, Andrew A., 1989. *Forecasting, Structural Time Series Models and The Kalman Filter*. Cambridge: Cambridge University Press.

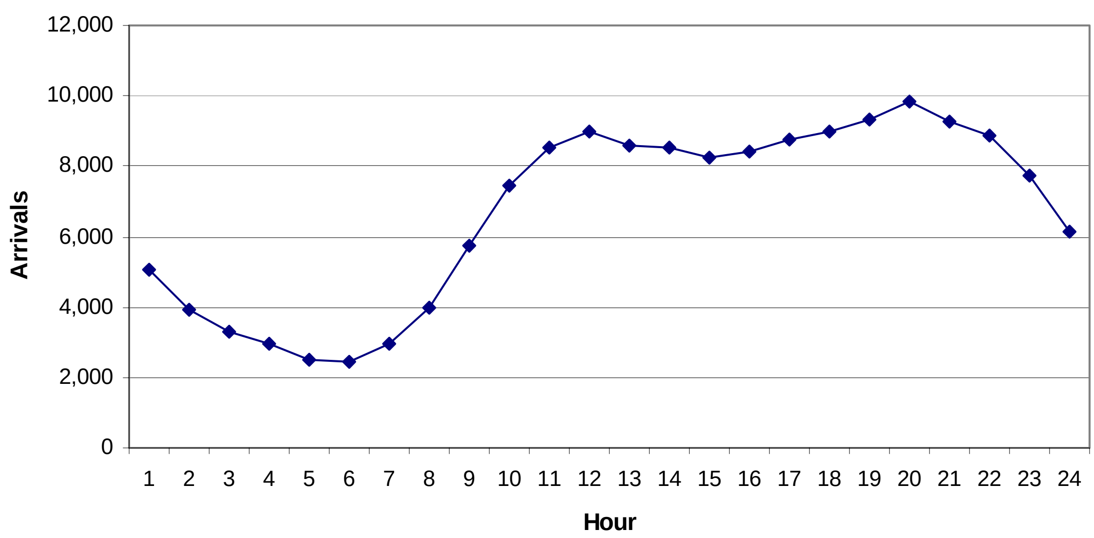
Mahapatra, S., L. Patvivastsiri, D. Eitel, C.P. Koelling, and L. Grove, 2004. "Does Clinical Re-Organization Through Simulation Hold The Key To Ed Re-Engineering Initiatives?" Paper and presentation at the Society For Health Systems, New Orleans.

Taylor, James W., L.M. de Menezes, and P.E. Mc Sharry, 2006. “A Comparison Of Univariate Methods For Forecasting Electricity Demand Up To A Day Ahead.” *International Journal of Forecasting*, 22, 1-16.

Table 1: Descriptive Statistics On Total Arrivals By Day Of The Week And Over The Entire Period

Statistic:	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Entire
Mean	185.52	170.61	159.34	159.42	158.11	158.28	177.74	166.98
Median	184.00	171.00	159.00	159.00	158.00	157.00	178.00	167.00
Mode	178.00	176.00	169.00	155.00	145.00	155.00	167.00	156.00
Standard deviation	17.34	14.75	15.33	14.21	14.61	12.78	12.86	17.85
Skewness	0.25	-0.05	-0.28	-0.08	0.25	-0.01	0.02	0.26
Kurtosis	0.51	-0.33	-0.55	-0.27	-0.29	0.83	0.07	0.24
Maximum	235.00	209.00	193.00	191.00	197.00	194.00	208.00	235.00
Minimum	140.00	132.00	121.00	126.00	124.00	117.00	144.00	117.00
Third quartile	196.00	182.00	171.00	169.00	169.00	167.00	184.00	178.00
First quartile	175.00	161.00	148.00	151.00	147.00	151.00	169.00	155.00

Figure 1: Arrivals by Hour



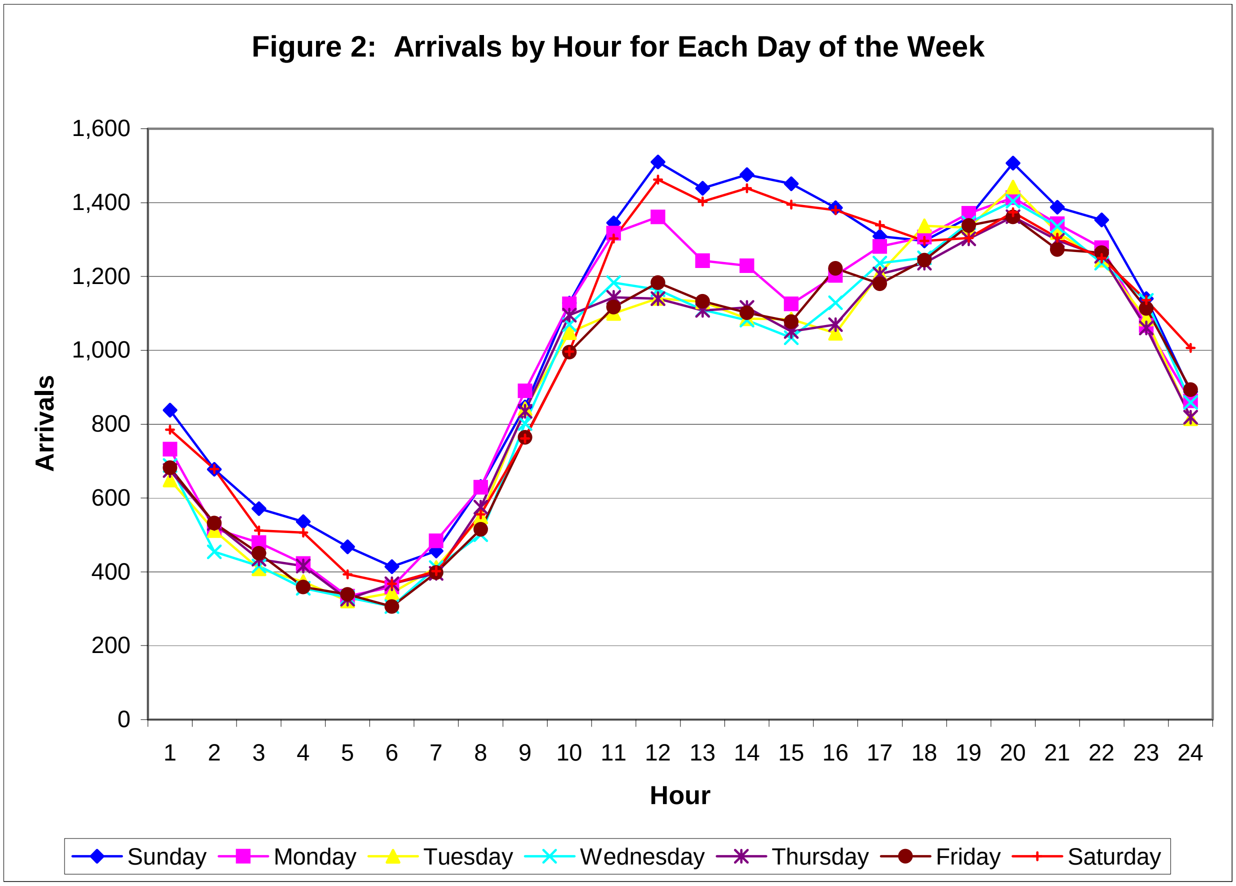


Figure 2: Arrivals by Hour for Each Day of the Week
Arrivals
1,600
1,400
1,200
1,000
800
600
400
200
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hour
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday

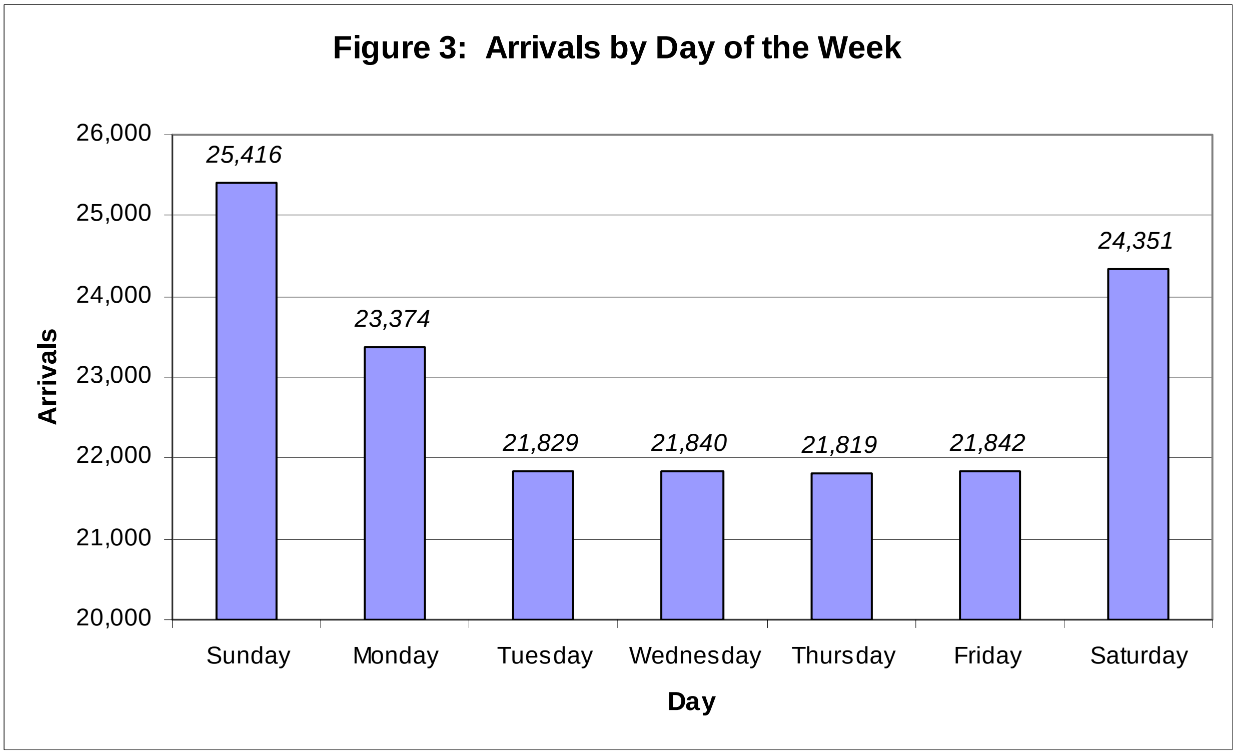


Figure 3: Arrivals by Day of the Week
Arrivals
26,000
25,000
24,000
23,000
22,000
21,000
20,000
25,416
23,374
21,829
21,840
21,819
21,842
24,351
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day

Figure 4: Daily Emergency Department Arrivals: November 11, 1999 to June 28, 2002

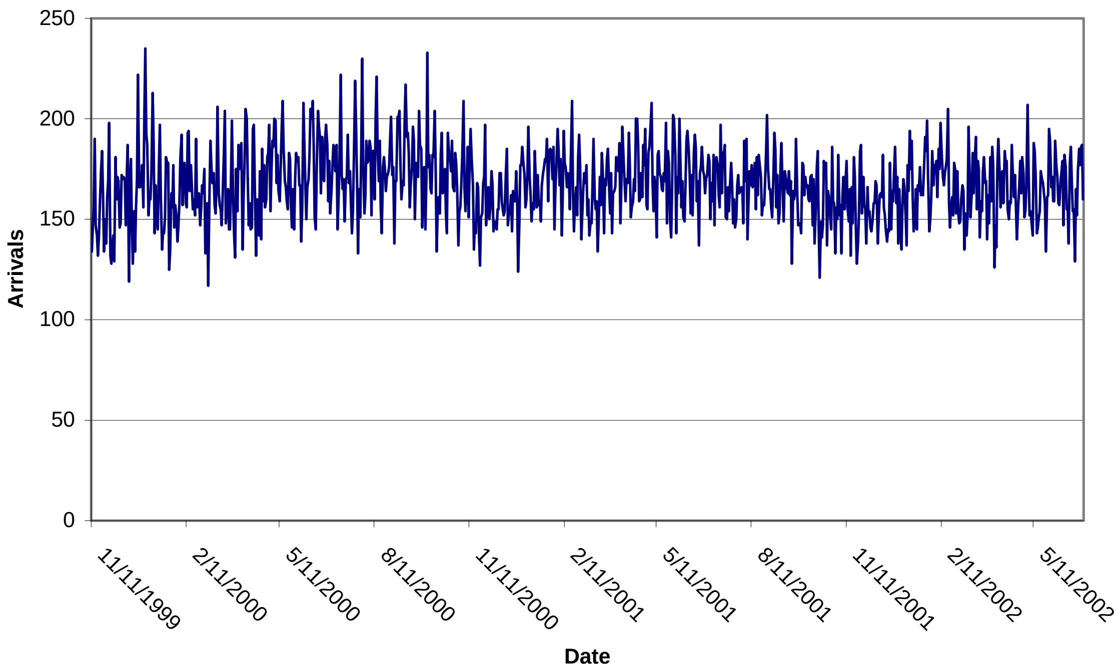


Figure 5: RMSE Comparisons, 1 To 168 Steps Ahead

Out-of-sample performance