

Statistics 531/Econ 677
Winter, 2007

We investigate over-crowding in the Emergency Room of the University of Michigan Hospital. The data, x_t , are hourly occupancy fractions for one year, starting July 1st 2005. Occupancy fraction is defined to be the mean number of patients in the ER during each hour divided by the total number of beds available (the ER operates 24 hours a day, 7 days a week, 365 days a year). Note that the occupancy fraction, shown in Fig. 1, can exceed one. The purposes of investigating these data are to predict future occupancy, and to make progress toward relating ER overcrowding with other variables such as errors in medical procedures.

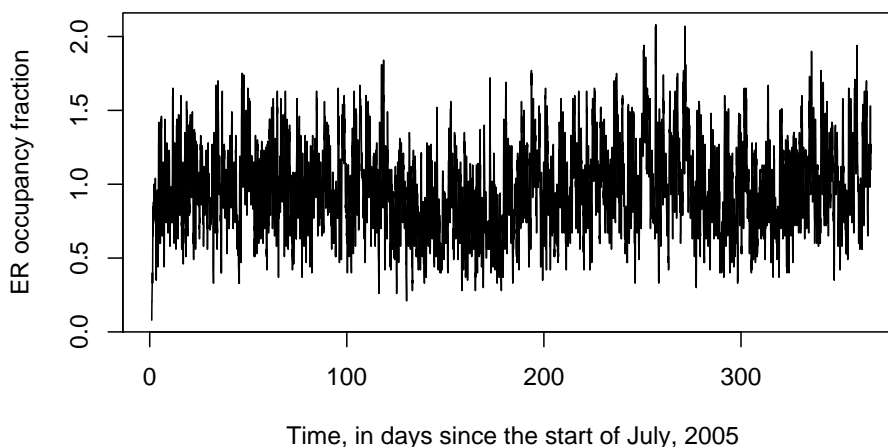


Figure 1: Hourly occupancy fraction at the University of Michigan Emergency Room

SECTION A. Fig. 2 shows a smoothed periodogram and an ACF of the data.

- A1. [1 point] What are the units of frequency in Fig. 2A? Explain your reasoning.
- A2. [2 points] Explain how you can tell that the periodogram in Fig. 2A has been truncated to exclude high frequencies (this is done to highlight the information about lower frequencies).
- A3. [3 points] Using Fig. 2, can you reject a null hypothesis that there is no weekly pattern to occupancy fraction? Explain.

SECTION B. Fig. 1 suggests that the occupancy could be modeled by a random process whose expected value μ_t is slowly varying with time. The variation around the mean in Fig. 1 appears quite stable. Thus, it may be reasonable to suppose that $x_t - \mu_t$ is stationary, where $\mu_t = E[x_t]$. We can estimate μ_t using local regression. This is done here using the R command `mu.hat=loess(x~ time,span=0.5)$fitted`. The estimate $\hat{\mu}_t$ of μ_t is shown in Fig. 3.

- B1. [2 points] Briefly describe what “local regression estimate” means.
- B2. [2 points] The dashed lines in Fig. 3 show an approximate 95% confidence interval, constructed by adding $\pm 2SE$ where SE is the standard error on the estimate of the mean, as calculated by the local regression. Is this interval appropriate? Explain. You should be able to comment even without a detailed understanding of local regression.

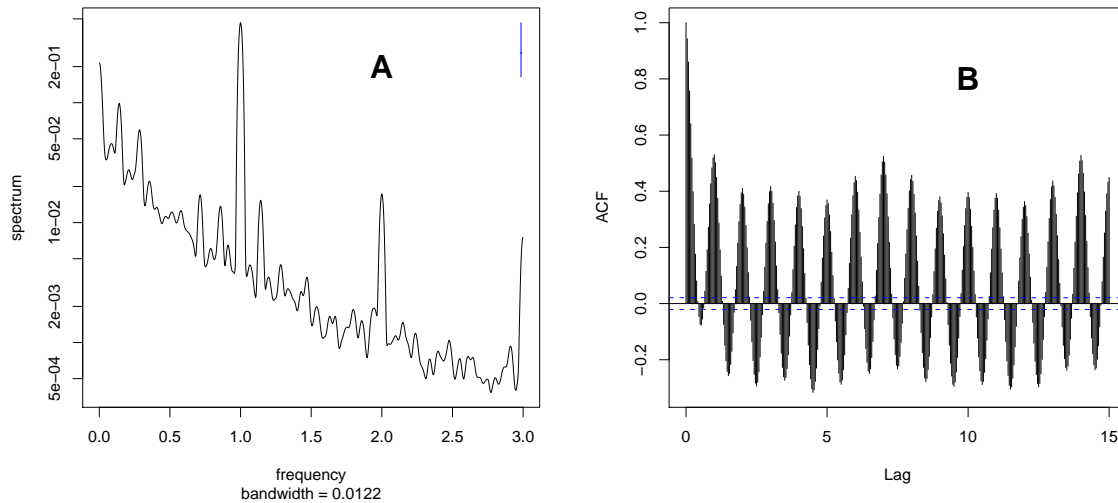


Figure 2: (A) Smoothed periodogram of x_t . (B) sample auto-covariance function of x_t

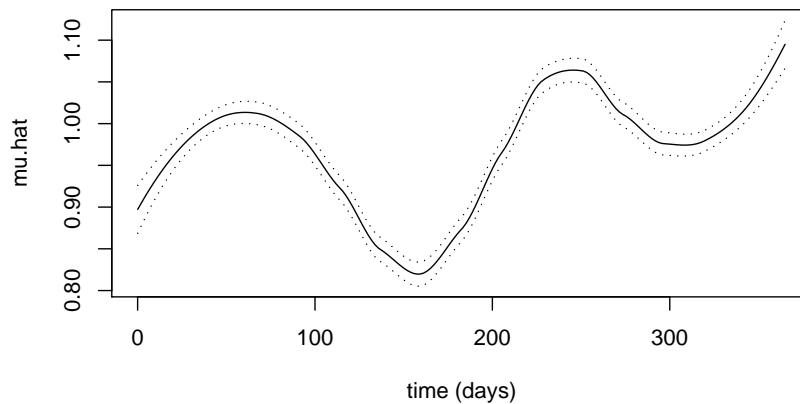


Figure 3: Estimate $\hat{\mu}_t$ of the mean hourly occupancy fraction μ_t .

B3. [2 points] Could the data be consistent with a model where the mean is not varying with time, e.g. a stationary process? Say yes or no, and explain.

SECTION C. We investigate the stationarity of the detrended occupancy fraction $y_t = x_t - \hat{\mu}_t$. In particular, we compare the two time intervals August/September 2005 and March/April 2006. First, we fit an $ARIMA(1,0,1) \times (1,0,1)_{24}$ model to the 61 days in August and September 2005. Below is the R output.

```

arima(x = y[AugSep], order = c(1, 0, 1), seasonal = c(1, 0, 1))
Coefficients:
      ar1      ma1      sar1      sma1  intercept
  0.9139  0.0403  0.9998 -0.9884   -0.0060
s.e.  0.0114  0.0277  0.0002  0.0080    0.1354
sigma^2 estimated as 0.006561:  log likelihood = 1568.48,  aic = -3124.96

```

C1. [3 points] Write out the fitted model, carefully stating all the model assumptions.

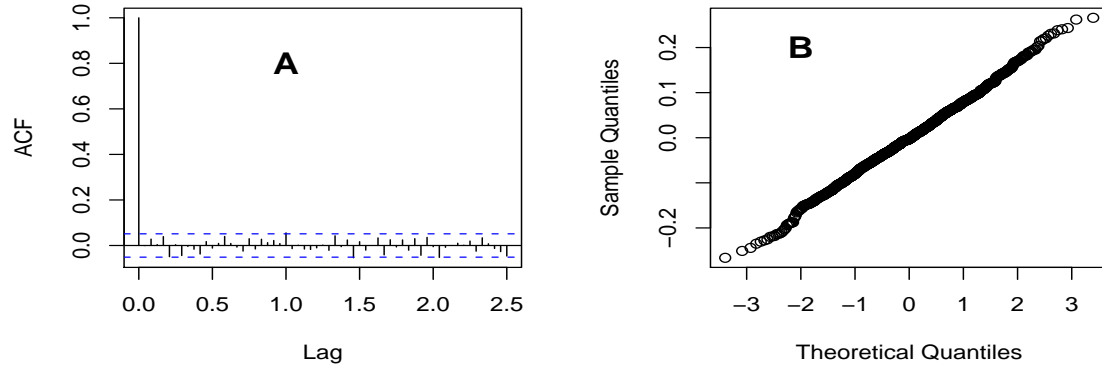


Figure 4: Investigation of the residuals from fitting an $ARIMA(1, 0, 1) \times (1, 0, 1)_{24}$ model to August/September. (A) Sample ACF. (B) normal quantile plot.

C2. [3 points] What do you conclude from the diagnostic plots in Fig. 4? Also, explain at least one relevant property that is NOT checked by these diagnostic plots, and describe how you could check it.

C3. [2 points] Based on Fig. 4 and the R model output above, $ARIMA(1, 0, 1) \times (0, 1, 1)_{24}$ was considered to be an acceptable model. Explain the reasoning.

C4. [2 points] A comparison of various models is presented in Table 1. What do you conclude from the AIC values in Table 1, together with the previous value of -3124.96 for $ARIMA(1, 0, 1) \times (1, 0, 1)_{24}$?

	$p \setminus q$	0	1	2		$p \setminus q$	0	1	2
A.	0	NA	-1612.9	-2258.7	B.	0	NA	-1168.5	-1844.2
	1	-3060.0	-3058.8	-3057.2		1	-2944.9	-2944.4	-2943.2
	2	-3058.8	-3057.1	-3055.2		2	-2944.5	-2943.1	-2941.3
	3	-3057.2	-3054.8	-3059.1		3	-2943.2	-2941.3	-2939.8

Table 1: AIC values from fitting $ARIMA(p, 0, q) \times (0, 1, 1)_{24}$ models to (A) August/September 2005, (B) March/April 2006.

Below is the R output from fitting an $ARIMA(1, 0, 0) \times (0, 1, 1)_{24}$ model to August/September 2005 and March/April 2006.

```

arima(x = y[AugSep], order = c(1, 0, 0), seasonal = c(0, 1, 1))
Coefficients:
      ar1      sma1
  0.9195  -1.0000
s.e.  0.0104  0.0197
sigma^2 estimated as 0.006496:  log likelihood = 1533,  aic = -3060

```

```

arima(x = y[MarApr], order = c(1, 0, 0), seasonal = c(0, 1, 1))

```

Coefficients:

	ar1	sma1
	0.9436	-1.0000
s.e.	0.0088	0.0341

sigma² estimated as 0.007036: log likelihood = 1475.46, aic = -2944.92

C5. [4 points] Show how to carry out an approximate hypothesis test that the AR1 component is the same for August/September 2005 and March/April 2006 in the context of an $ARIMA(1, 0, 0) \times (0, 1, 1)_{24}$ model. Explain what your approximations are for this test. How good do you think these approximations are, and how could you check? Note: since you are not provided with statistical tables, you are not required to calculate a p-value.