Analysis of U.S. Unemployment Rates

Outline

Monthly unemployment rates in Dallas County

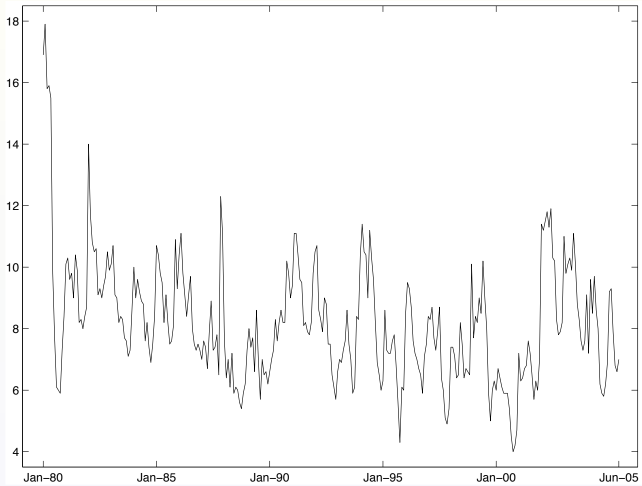

Figure 1: Monthly unemployment rates (in %) in Dallas County, Arizona, from January 1980 to June 2005. The data are obtained from the Website www.Economagic.com

ACFs and PACFs

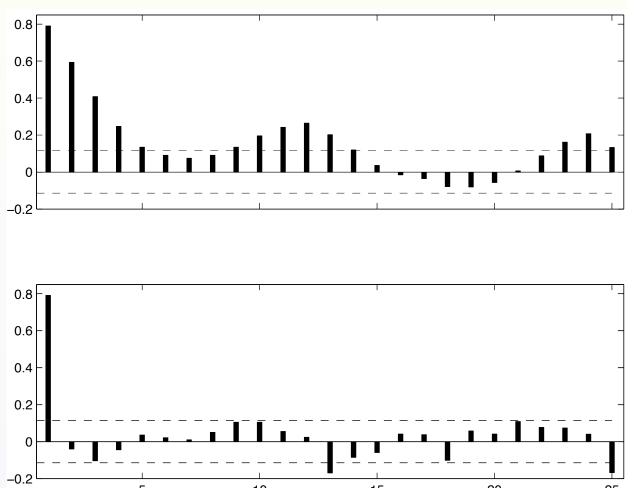

Figure 2: ACF (top panel) and PACF (bottom panel) of the unemployment rate. The dashed lines represent rejection boundaries of 5%-level tests of zero ACF and PACF at indicated lag.

Splitting the time series

- We split the time series into a training sample of historical data from January 1980 to December 2004 and a second sample of "test data" from January to June 2005. The second sample is used to measure the performance of the out-of-sample forecasts developed from the training sample.
- Since there are obvious seasonal effects on unemployment, we use the `R` or `S` function `stl` to decompose the training sample into a trend, a seasonal component, and residual; Figure 3 plots the trend and seasonal components and the residuals of the decomposition, with a period of 12 months for the seasonal component.

Splitting the time series

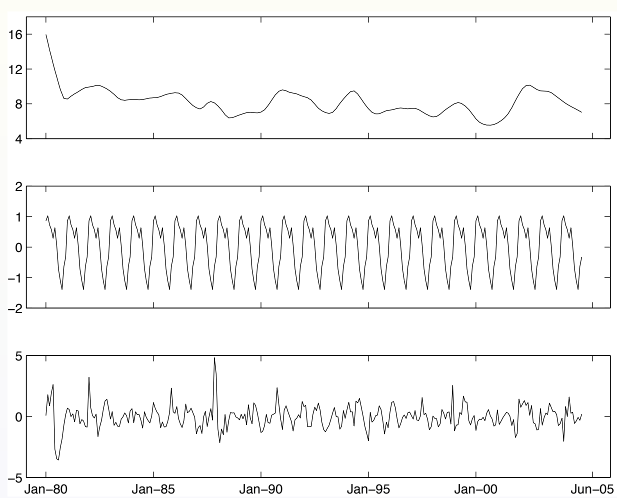

Figure 3: Decomposition of the time series of unemployment rates into the trend (top panel), the seasonal component (middle panel), and residuals (bottom panel).

ACFs and PACFs of the training sample

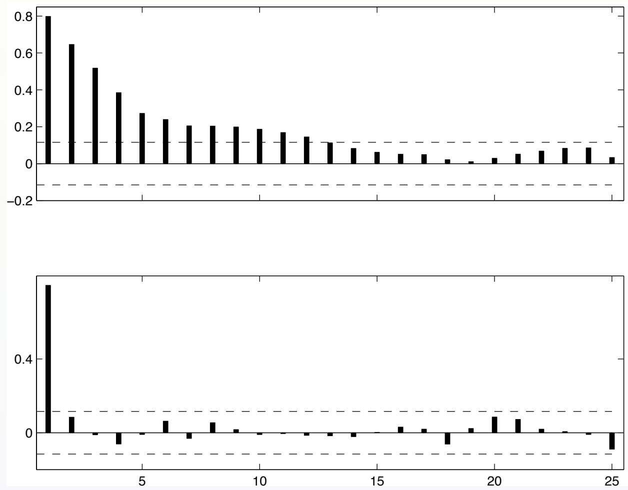

Figure 4: ACF (top panel) and PACF (bottom panel) of the deseasonalized time series from the training sample. The dashed lines represent rejection boundaries of 5%-level tests of zero ACF and PACF at indicated lag.

ARMA(1,1) estimation

- We fit an ARMA model to the deseasonalized series by using the AIC to determine the order of the ARMA model. Based on the AIC, we fit the ARMA(1,1) model

$$\begin{aligned} x_t &= \underset{(0.4417)}{8.3421} + \underset{(0.0550)}{0.8828}x_{t-1} + \epsilon_t - \underset{(0.067)}{0.099}\epsilon_{t-1}, \quad \epsilon_t \sim N(0,\sigma^2) \end{aligned}$$

  with the standard errors of the parameter estimates given in parentheses. The results are obtained by using the `R` function `arima`, which also gives $\widehat{\sigma}^2 = 1.025$.
- Most of the standardized residuals $\epsilon_t/\sqrt{\text{Var}(\epsilon_t)}$ are small, and so are their autocorrelations shown in the top panel of Figure 5.
- Advanced material: The bottom panel of Figure 5 shows the p-values of the Ljung-Box statistics $Q(m)$ for different values of the lag m. These p-values are well above the level 0.05 (shown by the broken line), below which the Ljung-Box test rejects the null hypothesis of zero autocorrelations up to lag m.

Residual analysis

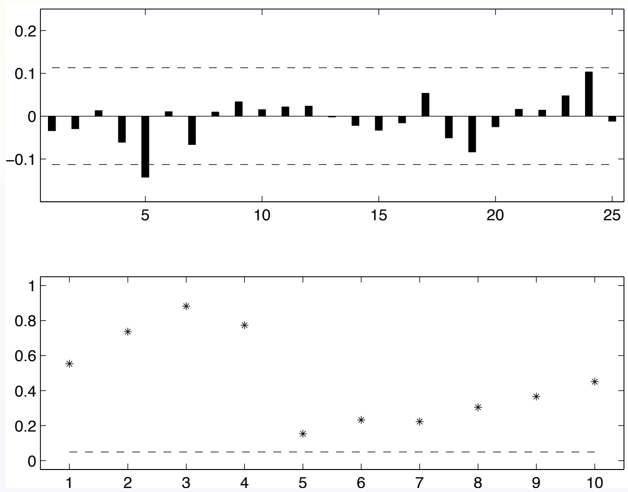

Figure 5: Diagnostic plots for the fitted ARMA(1, 1) model. Top panel: ACF of residuals; bottom panel: p-values of Ljung-Box statistics $Q(m)$.

Forecasting

Table 1: Forecasts of unemployment rates in 2005.

	Jan.	Feb.	Mar.	Apr.	May	June
Actual rate	9.2	9.3	7.9	6.8	6.6	7.0
Deseasonalized rate	7.21	7.48	7.65	7.79	7.90	7.98
(s.e.)	(1.14)	(1.49)	(1.75)	(1.89)	(1.98)	(2.02)
Seasonal rate	0.86	1.03	0.73	0.57	0.29	0.63
Predicted rate	8.07	8.51	8.38	8.36	8.19	8.61

We can use the fitted ARMA(1, 1) model to obtain k-months-ahead forecasts. The `R` function `arima` can be used to calculate these forecasts and their standard errors (s.e.). Table 1 gives the forecast values of the deseasonalized series from January to June 2005 based on the ARMA(1,1) model fitted to the training sample from January 1980 to December 2004.

R code of the analysis

```
### Analyze Unemployment Rate in Dallas County, AR: Percent: NSA
### (Jan-1980 -- Jun -2005), Data after Dec, 2004 are used for comparing
### with the prediction results

> series<-read.table("http://www.stanford.edu/~xing/statfinbook/_BookData
/Chap05/unem_dallas.txt", skip=1)
> unem<-ts(series[1:300,3], freq=12, start=c(1980, 1))
> ts.plot(unem)

### Plot acf and pacf
> par(mfrow=c(2,1)); acf(unem); pacf(unem)

## Use 'stl' to decompose the series, and leave out the seasonal components
> unem.stl <- stl(unem, "periodic")   ## seasonal, trend, resid
> unem.series <-unem.stl$time[,2]+unem.stl$time[,3]
```


Data Estimation Forecasting

R code of the analysis

```
### Fit ARMA(1,1) model to data
> library(MASS)
> (unem.series.arma1<-arima(unem.series, order=c(1,0,1)))
Call:
arima(x = unem.series, order = c(1, 0, 1))

Coefficients:
        ar1    ma1  intercept
     0.8828 -0.099     8.3421
s.e. 0.0550  0.067     0.4417

sigma^2 estimated as 1.025:  log likelihood = -430.05,  aic = 868.09
```

AMS316, Stony Brook University

R code of the analysis

```
### Prediction of the series
> unem.pred<-predict(unem.series.arma1, n.ahead=6)

> unem.pred
$pred
          Jan      Feb      Mar      Apr      May      Jun
2005 7.340738 7.459020 7.563442 7.655627 7.737010 7.808856
$se
          Jan      Feb      Mar      Apr      May      Jun
2005 1.012390 1.286331 1.464726 1.589943 1.681077 1.748814

> unem.pred$pre+unem.sea[1:6]
          Jan      Feb      Mar      Apr      May      Jun
2005 8.202830 8.485186 8.293680 8.226127 8.027772 8.440197

> ts(series[301:306,3], freq=12, start=c(2005,1))
     Jan Feb Mar Apr May Jun
2005 9.2 9.3 7.9 6.8 6.6 7.0
```