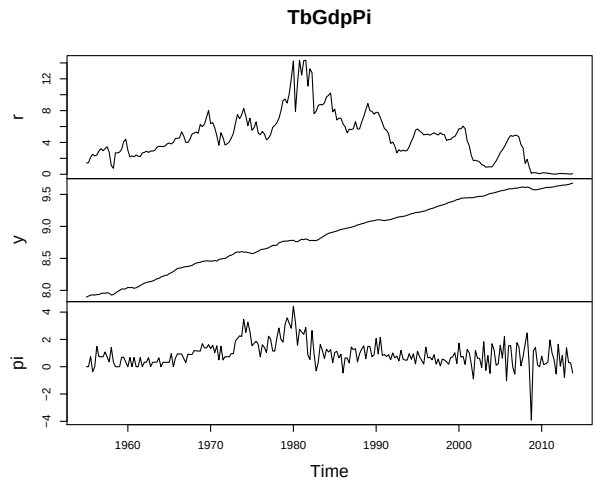
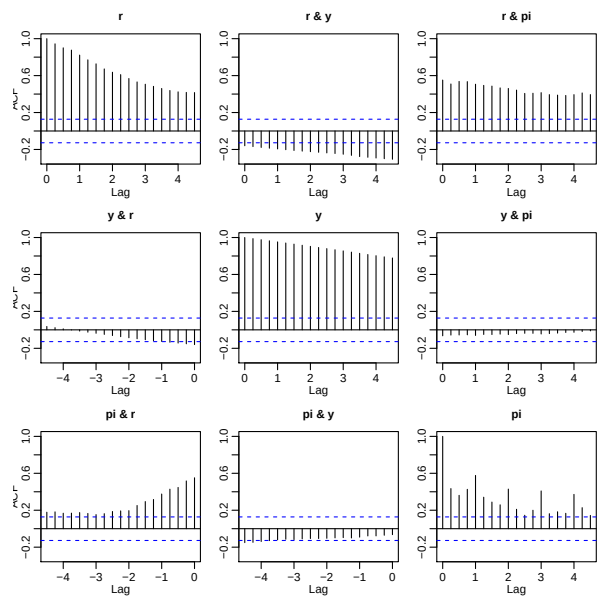
Solutions to Selected Computer Lab Problems and Exercises in Chapter 12 of *Statistics and Data Analysis for Financial Engineering, 2nd ed.* by David Ruppert and David S. Matteson





Problem 1. (a) None of the three times series plots show clear signs of being mean-reverting. The plot of the log of GDP is clearly non-stationary and the Treasury rate and inflation rate may also be non-stationary.

In the ACF plots we see slow nearly linear decays, especially for log(GDP) which also suggests that none of the series are stationary.

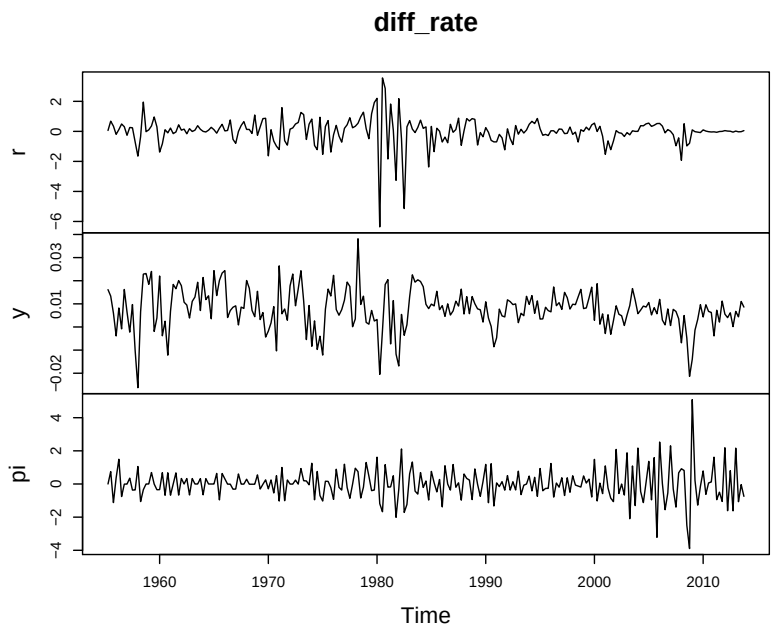
(b)

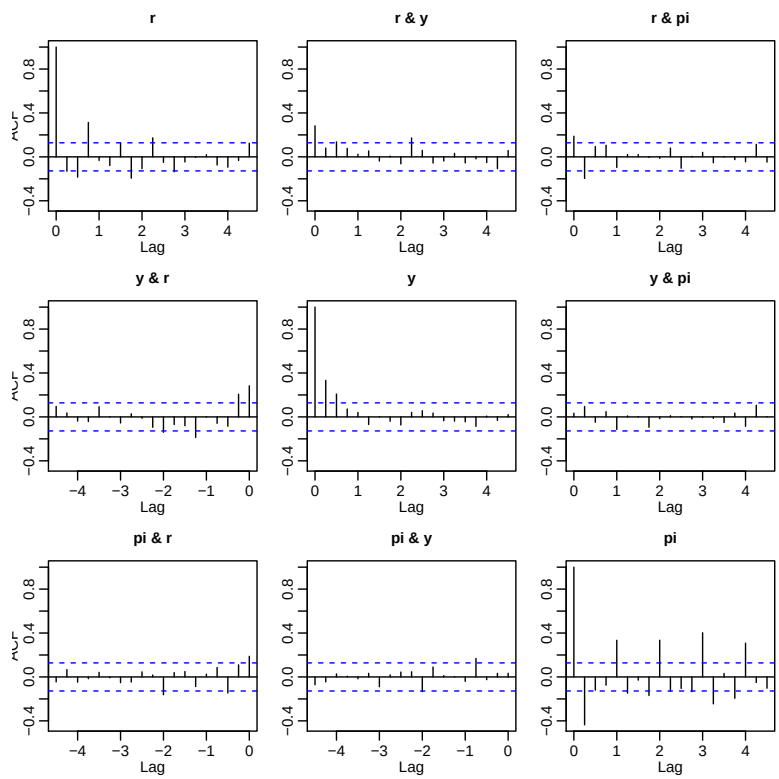
In the ADF tests, the null hypothesis is unit-root nonstationarity and the alternative is stationary. None of the p-values is below 0.05, and two are quite large, so we probably do not want to reject the null hypotheses that three series are nonstationary. The ADF tests corroborate the conclusions from the plots that all three series are non-stationary.

```
> adf.test(TbGdpPi[,1])

Augmented Dickey-Fuller Test

data:  TbGdpPi[, 1]
Dickey-Fuller = -2.5622, Lag order = 6, p-value = 0.3395
alternative hypothesis: stationary

> adf.test(TbGdpPi[,2])

Augmented Dickey-Fuller Test

data:  TbGdpPi[, 2]
Dickey-Fuller = -1.3474, Lag order = 6, p-value = 0.8502
alternative hypothesis: stationary

> adf.test(TbGdpPi[,3])

Augmented Dickey-Fuller Test

data:  TbGdpPi[, 3]
Dickey-Fuller = -3.1444, Lag order = 6, p-value = 0.09794
alternative hypothesis: stationary
```

Problem 2. 1. The differenced series appear stationary according to the augmented Dickey-Fuller tests since all p-values are smaller than 0.01, so we reject the null hypothesis of nonstationarity for each series. Also, the time series plots show signs of mean-reversion and the ACF plots show rapid decay to 0, except that the differenced inflation rate series shows persistent autocorrelation at lags that are a multiple of one year.

2. All three series show autocorrelation. The differenced Treasury bill series shows an irregular pattern of autocorrelation. The differenced log real GDP series has autocorrelation that decays to zero much like an AR(1) process. As mention in 1., the differenced inflation rate series shows autocorrelation at yearly lags.

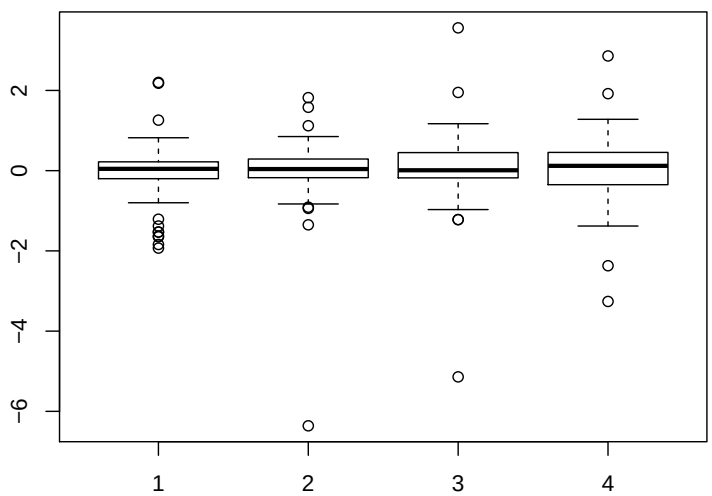





```
> diff_rate = diff(TbGdpPi)
> adf.test(diff_rate[,1])

Augmented Dickey-Fuller Test

data:  diff_rate[, 1]
Dickey-Fuller = -6.3425, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
```

```
In adf.test(diff_rate[, 1]) : p-value smaller than printed p-value
> adf.test(diff_rate[,2])

Augmented Dickey-Fuller Test

data:  diff_rate[, 2]
Dickey-Fuller = -5.6647, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diff_rate[, 2]) : p-value smaller than printed p-value
> adf.test(diff_rate[,3])

Augmented Dickey-Fuller Test

data:  diff_rate[, 3]
Dickey-Fuller = -7.9521, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diff_rate[, 3]) : p-value smaller than printed p-value
```

Problem 3. The boxplots do not show any seasonal differences, since the four boxplots are similar to each other. Although you were not asked to perform a formal test of differences, the one-way ANOVA test below fails to reject the null hypothesis that the four seasons have the same mean.



These results justify using a nonseasonal time series model for the T-bill series.

```
> anova(lm(diff_rate[,1] ~ factor(cycle(diff_rate))))
Analysis of Variance Table

Response: diff_rate[, 1]
                          Df  Sum Sq Mean Sq F value Pr(>F)
factor(cycle(diff_rate))   3   0.382 0.12748   0.146 0.9322
Residuals                231 201.659 0.87298
```

Problem 4. 1. First differencing, that is, $d = 1$.

2. ARIMA(2,1,3)

3. AIC. (Part 2. gives this away.)

```
> auto.arima(TbGdpPi[,1], max.P=0, max.Q=0, ic="aic")
Series: TbGdpPi[, 1]
ARIMA(2,1,3)

Coefficients:
         ar1      ar2      ma1     ma2     ma3
      0.1368  -0.2225  -0.2597  0.1041  0.3364
s.e.  0.1655   0.1483   0.1539  0.1483  0.0649

sigma^2 estimated as 0.7418:  log likelihood=-298.6
AIC=609.19   AICc=609.56   BIC=629.95
```

4. Yes there is a change and the best-fitting model is now ARIMA(2,1,2).

```
> auto.arima(TbGdpPi[,1], max.P=0, max.Q=0, ic="bic")
Series: TbGdpPi[, 1]
ARIMA(2,1,2)

Coefficients:
          ar1      ar2     ma1     ma2
      -0.8999  -0.7403  0.8139  0.4912
s.e.   0.1094   0.1079  0.1420  0.1419

sigma^2 estimated as 0.7488:  log likelihood=-299.64
AIC=609.27   AICc=609.53   BIC=626.57
```

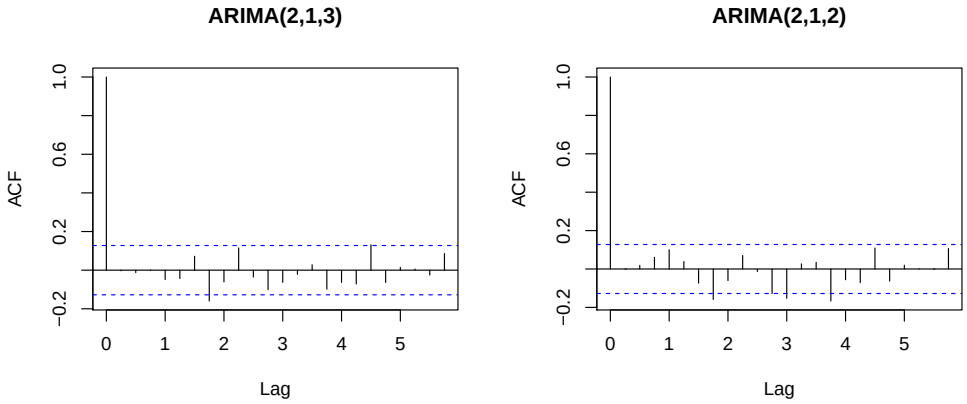
5.

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

where $\Delta Y_t = Y_t - Y_{t-1}$ for every t, $\epsilon_1, \ldots$ is a white noise process, and ϕ_1, ϕ_2, θ_1, and θ_2 are parameters whose estimates are -0.8999, -0.7403, 0.8139, and 0.4912, respectively. There is no μ in this model; the mean is 0.

Problem 5. Since the best-fitting model differ between the two criteria, the results below are for both fits. We see that there is statistically significant residual autocorrelation. However, none of the residual autocorrelations are much larger than 0.2 in magnitude and the larger residual autocorrelations occur at lags of at least 7 quarters. Therefore, we might be satisfied with either of these models, since the residual autocorrelations at short lags are extremely small.

If we wished to consider other types of models, autocorrelations would be reduced by a nonseasonal ARIMA model only if that model had a very large number of parameters. Seasonal ARIMA models are discussed in Section 13.1, but these do not reduce the residual autocorrelations much. Another possible model would be a long-memory process; these are discussed in Section 13.5 and a different inflation rate series, the monthly series used in Example 12.1, is used as an example.




```
> fit1 = arima(TbGdpPi[,1], order=c(2,1,3))
> pdf("residACF.pdf",width=9,height =4)
> par(mfrow=c(1,2))
> acf(residuals(fit1),main = "ARIMA(2,1,3)")
> Box.test(residuals(fit1), lag = 12, type="Ljung", fitdf=5)

Box-Ljung test

data:  residuals(fit1)
X-squared = 16.643, df = 7, p-value = 0.01985

>
> fit2 = arima(TbGdpPi[,1], order=c(2,1,2))
> acf(residuals(fit2),main = "ARIMA(2,1,2)")
> Box.test(residuals(fit2), lag = 12, type="Ljung", fitdf=4)

Box-Ljung test

data:  residuals(fit2)
X-squared = 23.33, df = 8, p-value = 0.002966
```

Exercise 1.

$$\widehat{Y}_{n+1} = 104 + 0.4 * (99 - 104) + 0.25 * (103 - 104) + 0.1 * (102 - 104) = 102.$$

$$\widehat{Y}_{n+2} = 104 + 0.4 * (102 - 104) + 0.25 * (99 - 104) + 0.1 * (103 - 104) = 102.$$

Exercise 7. (a) To simplify notation assume that $\mu = 0$ or, equivalently, that in the following calculations μ has been subtracted from the process. Then

$$\gamma(k) = E(Y_{t+k}Y_t) = E\left\{(\phi_1 Y_{t+k-1} + \phi_2 Y_{t+k-2} + \epsilon_{t+k})Y_t\right\} = \phi_1\gamma(k-1) + \phi_2\gamma(k-2).$$

The result follows by dividing through by $\gamma(0)$.

(b) The result in part (a) with $k = 1$ is

$$\rho(1) = \phi_1 + \phi_2\rho(1),$$

since $\rho(0) = 1$ and $\rho(-1) = \rho(1)$. The result with $k = 2$ is

$$\rho(2) = \phi_1\rho(1) + \phi_2,$$

again using $\rho(0) = 1$. Putting these two equations in matrix form finishes the exercise.

(c) $\widehat{\phi}_1 = 0.3810$ and $\widehat{\phi}_2 = 0.0476$. Also, $\widehat{\rho}(3) = 0.0952$. See the `R` calculations below.

```
> x = solve(matrix(c(1,0.4,0.4,1),nrow=2))
> phi12 = x %*% c(0.4,0.2)
> options(digits=3)
> phi12
       [,1]
[1,] 0.3810
[2,] 0.0476
> c(0.2,0.4) %*% phi12
       [,1]
[1,] 0.0952
```

Exercise 17. A polynomial trend means that $\mu(t) = \sum_{j=0}^{m} \beta_j t^j$ for some nonnegative integer m; this definition includes the degenerate case $m = 0$ where the mean is constant. Since,

$$\Delta(t^p) = t^p - (t-1)^p = t^p - \sum_{j=0}^{p} \binom{p}{j} t^j(-1)^{p-j} = -\sum_{j=0}^{p-1} \binom{p}{j} t^j(-1)^{p-j}$$

we see that if $P(t)$ is a polynomial of degree m, then $\Delta\{P(t)\}$ is a polynomial of degree $p-1$ and $\Delta^d(P(t))$ is a polynomial of degree $m-d$ for $d \leq m$. In the case $d = m$, $\Delta^d\{P(t)\}$ is a constant (a zero-degree polynomial). Since $\Delta c = c - c = 0$ for any constant c, $\Delta^d\{P(t)\} = 0$ for $d > m$.