**THE UNIVERSITY OF CHICAGO**
**Booth School of Business**
Business 41914, Spring Quarter 2015, Mr. Ruey S. Tsay

**Lecture 8: Seasonal Model, Principal Component Analysis and Factor Models**

**Reference**: Chapter 6 of the textbook. Data sets used are available from text web.

## 6.1 Seasonal Models

A direct generalization of the useful univariate *airline* model for the vector seasonal time series $\boldsymbol{z}_t$ is

$$(1-B)(1-B^s)\boldsymbol{z}_t = (\boldsymbol{I}_k - \boldsymbol{\theta}B)(\boldsymbol{I}_k - \boldsymbol{\Theta}B^s)\boldsymbol{a}_t, \tag{6.1}$$

where all eigenvalues of $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ are less than 1 in modulus, $s > 1$ denotes the number of seasons within a year, and $\{\boldsymbol{a}_t\}$ is a sequence of iid random vectors with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}_a$. For instance, $s = 4$ for quarterly data and $s = 12$ for monthly data. For the model in Equation (6.1), $(1 - B^s)$ is referred to as the seasonal difference. We refer to the model as the seasonal airline model. Rewriting Equation (6.1) as

$$\boldsymbol{z}_t = \frac{\boldsymbol{I}_k - \boldsymbol{\theta}B}{1-B} \times \frac{\boldsymbol{I}_k - \boldsymbol{\Theta}B^s}{1-B^s}\boldsymbol{a}_t,$$

we see that the model continues to represent double exponential smoothing with one smoothing for the regular dependence and anther one for the seasonal dependence.
Let $\boldsymbol{w}_t = (1-B)(1-B^s)\boldsymbol{z}_t$ and $\boldsymbol{\Gamma}_{w,\ell}$ be the lag-$\ell$ auto-covariance matrix of $\boldsymbol{w}_t$. It is straightforward to see that $\boldsymbol{w}_t = (\boldsymbol{I}_k - \boldsymbol{\theta}B - \boldsymbol{\Theta}B^s + \boldsymbol{\theta}\boldsymbol{\Theta}B^{s+1})\boldsymbol{a}_t$ and

1. $\boldsymbol{\Gamma}_{w,0} = (\boldsymbol{\Sigma}_a + \boldsymbol{\Theta}\boldsymbol{\Sigma}_a\boldsymbol{\Theta}') + \boldsymbol{\theta}(\boldsymbol{\Sigma}_a + \boldsymbol{\Theta}\boldsymbol{\Sigma}_a\boldsymbol{\Theta}')\boldsymbol{\theta}'$,
2. $\boldsymbol{\Gamma}_{w,1} = -\boldsymbol{\theta}(\boldsymbol{\Sigma}_a + \boldsymbol{\Theta}\boldsymbol{\Sigma}_a\boldsymbol{\Theta}')$,
3. $\boldsymbol{\Gamma}_{w,s-1} = \boldsymbol{\Theta}\boldsymbol{\Sigma}_a\boldsymbol{\theta}'$,
4. $\boldsymbol{\Gamma}_{w,s} = -(\boldsymbol{\Theta}\boldsymbol{\Sigma}_a + \boldsymbol{\theta}\boldsymbol{\Theta}\boldsymbol{\Sigma}_a\boldsymbol{\theta}')$,
5. $\boldsymbol{\Gamma}_{w,s+1} = \boldsymbol{\theta}\boldsymbol{\Theta}\boldsymbol{\Sigma}_a$,
6. $\boldsymbol{\Gamma}_{w,\ell} = \boldsymbol{0}$, otherwise.

Since matrix multiplication is generally not commutable, further simplification in the formulas for $\boldsymbol{\Gamma}_\ell$ is not always possible. Let $\boldsymbol{\rho}_{w,\ell}$ be the lag-$\ell$ cross-correlation matrix of $\boldsymbol{w}_t$. The above results show that the $\boldsymbol{w}_t$ series has non-zero cross-correlation matrices at lags 1, $s-1$, $s$, and $s+1$. Thus, $\boldsymbol{w}_t$ has the same non-zero lag dependence as its univariate counterpart, but it does not share the same symmetry in correlations, because $\boldsymbol{\rho}_{w,s-1} \neq \boldsymbol{\rho}_{w,s+1}$ in general.
In some applications, the seasonal airline model can be generalized to

$$(1-B)(1-B^s)\boldsymbol{z}_t = \boldsymbol{\theta}(B)\boldsymbol{\Theta}(B)\boldsymbol{a}_t, \tag{6.2}$$

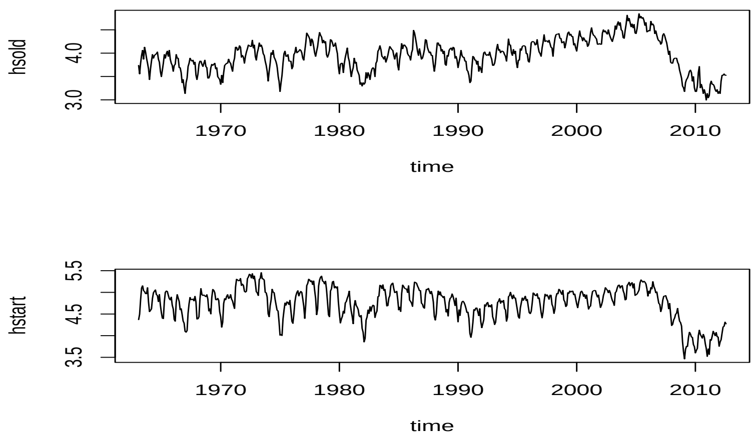


Figure 6.1: Time plots of monthly U.S. housing data from January 1963 to July 2012: (a) upper plot: new homes sold and (b) new privately owned housing units started. Both series are logged.

where both $\boldsymbol{\theta}(B)$ and $\boldsymbol{\Theta}(B)$ are matrix polynomials of order $q$ and $Q$, respectively, with $q < s$. Limited experience indicates that $Q = 1$ is often sufficient in applications.

**Example 6.1**. Consider the monthly housing data of the U.S. from January 1963 to July 2012. The two series employed are

1. $z_{1t}$: Logarithm of new homes sold in thousands of units (new residential sales),

2. $z_{2t}$: Logarithm of the total new privately owned housing units started in thousands of units (new residential construction).

Figure 6.1 shows the time plots of the two logged time series. A strong seasonal pattern is clearly seen from the plots. The impact of 2007 sub-prime financial crisis and the subsequent prolong decline in the housing market are also visible.
Figure 6.2 shows the cross-correlation matrices of $\boldsymbol{w}_t = (1-B)(1-B^{12})\boldsymbol{z}_t$. From the plots, the main non-zero cross correlation matrices are at lags 1, 11, 12, and 13 with some additional minor cross dependence at lags 2 and 3. These plots give some justifications for employing a multiplicative seasonal model. To begin with, we consider the univariate seasonal models and obtain

$$
\begin{aligned}
w_{1t} &= (1-0.21B)(1-0.90B^{12})a_{1t} && (6.3)\\
w_{2t} &= (1-0.33B+0.04B^2+0.11B^3)(1-0.86B^{12})a_{2t}, && (6.4)
\end{aligned}
$$

where $\boldsymbol{w}_t = (1-B)(1-B^{12})\boldsymbol{z}_t$, and the residual variances are 0.0069 and 0.0080, respectively, for $a_{1t}$ and $a_{2t}$. Both estimates in Equation (6.3) are statistically significant at the usual 5% level. The standard errors of the regular MA coefficients in Equation (6.4) are 0.042, 0.042, and 0.041, respectively, implying that the lag-2 coefficient is not statistically significant at the conventional 5% level.

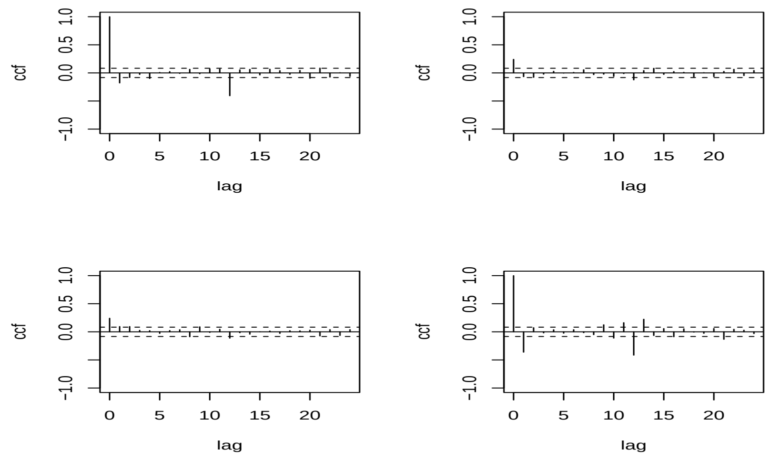

Figure 6.2: Sample cross-correlation coefficients of $\boldsymbol{w}_t = (1-B)(1-B^{12})\boldsymbol{z}_t$ for the U.S. logged housing data.

Turn to multivariate seasonal analysis. Based on the cross-correlation matrices in Figure 6.2 and the results of univariate models, we employ the model

$$(1-B)(1-B^{12})\boldsymbol{z}_t = (\boldsymbol{I}_k - \boldsymbol{\theta}_1 B - \boldsymbol{\theta}_2 B^2 - \boldsymbol{\theta}_3 B^3)(\boldsymbol{I}_k - \boldsymbol{\Theta} B^{12})\boldsymbol{a}_t. \tag{6.5}$$

The parameter estimates and those of a refined version of the model in Equation (6.5) are given in Table 6.1. The refinement is to remove some insignificant parameters. Both AIC and BIC prefer the refined model over the unconstrained one. Standard errors of the coefficient estimates are given in the attached R demonstration. From the table, we see that large coefficients appear in $\boldsymbol{\theta}_1$ and $\boldsymbol{\Theta}$, indicating that the fitted models are close to that of a multivariate airline model. Figure 6.3 shows the time plots of the residuals of the reduced model in Table 6.1 whereas Figure 6.4 plots the cross-correlation matrices of the residuals. The fitted model appears to be adequate as there is no major significant cross correlations in the residuals.
Comparing with the univariate seasonal models, we make the following observations. (a) The model for $z_{1t}$ (houses sold) in the VARMA model is to be close to the univariate model in Equation (6.3). For instance, the residual variances of the two models are close and elements of $\widehat{\boldsymbol{\theta}}_3$ are small. (b) The VARMA model seems to provide a better fit for $z_{2t}$ (housing starts) because the VARMA model has a smaller residual variance. (c) The VARMA model shows that there exists a feedback relationship between houses sold and housing starts. The cross dependence is strong at lag 1, but only marginal at the seasonal lag. Finally, predictions and forecast error variance decomposition of the fitted seasonal model can be carried out in the same way as that of the non-seasonal VARMA models. Details are omitted.

**Remark**: The commands used to estimate seasonal VARMA models in the `MTS` package are `sVARMA` and `refsVARMA`. These commands use the conditional maximum likelihood method. The model is

Table 6.1: Estimation Results of Seasonal Models for Monthly Housing Data

| Parameter | Full Model | | Reduced Model | | Switched | |
|---|---|---|---|---|---|---|
| $\boldsymbol{\theta}_1$ | 0.246 | -0.168 | 0.253 | -0.168 | 0.246 | -0.168 |
| | -0.290 | 0.416 | -0.292 | 0.416 | -0.297 | 0.416 |
| $\boldsymbol{\theta}_2$ | 0.028 | 0.034 | 0 | 0 | 0.022 | 0.040 |
| | -0.168 | 0.048 | -0.168 | 0 | -0.168 | 0.048 |
| $\boldsymbol{\theta}_3$ | 0.063 | -0.067 | 0.076 | -0.068 | 0.057 | -0.062 |
| | -0.085 | -0.074 | -0.085 | -0.073 | -0.085 | -0.068 |
| $\boldsymbol{\Theta}$ | 0.850 | 0.033 | 0.850 | 0.033 | 0.850 | 0.037 |
| | 0.044 | 0.824 | 0.044 | 0.824 | 0.029 | 0.848 |
| $10^2 \times \boldsymbol{\Sigma}_a$ | 0.694 | 0.263 | 0.696 | 0.266 | 0.694 | 0.261 |
| | 0.263 | 0.712 | 0.266 | 0.714 | 0.261 | 0.712 |
| AIC & BIC | -10.012 | -9.892 | -10.017 | -9.920 | -10.009 | -9.889 |

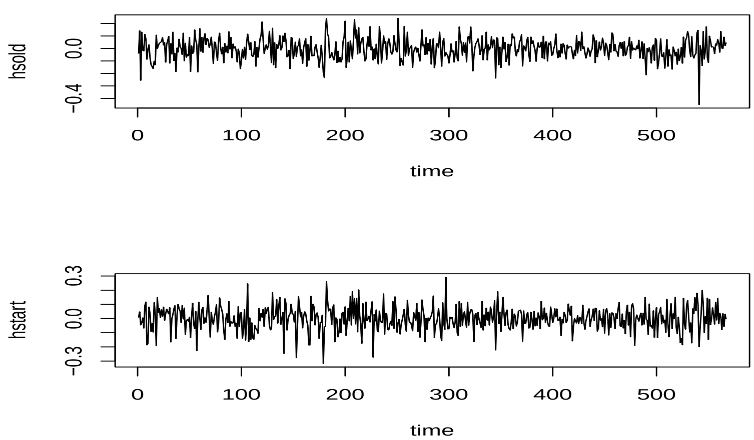


Figure 6.3: Residuals of the reduced model in Table 6.1 for the monthly housing data in logarithms.

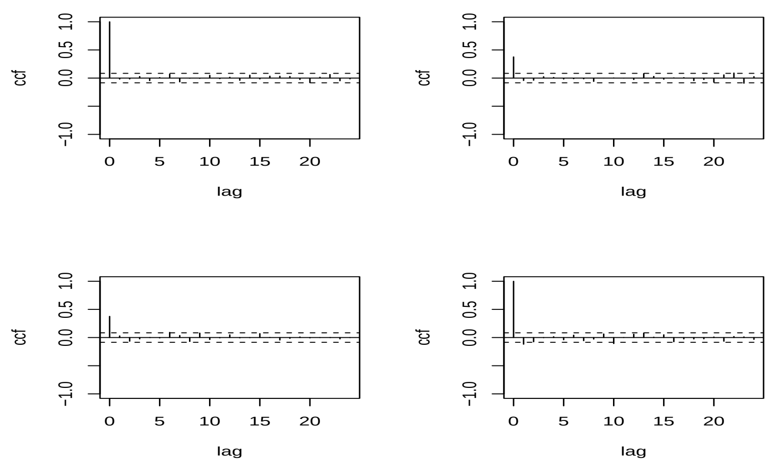


Figure 6.4: Cross-correlations of the residuals of the reduced model in Table 6.1 for the U.S. monthly housing data.

specified by the regular order $(p, d, q)$ and seasonal order $(P, D, Q)$ with $s$ denoting the seasonality.

***R Demonstration***: Estimation of Seasonal VARMA models. Output edited.

```
> m3=sVARMA(zt,order=c(0,1,3),sorder=c(0,1,1),s=12)
Number of parameters:  16
initial estimates:  0.1985901 -0.08082791 0.05693362 ....
Coefficient(s):
       Estimate  Std. Error  t value Pr(>|t|)
 [1,]   0.24587     0.04621    5.321 1.03e-07 ***
 [2,]  -0.16847     0.04483   -3.758 0.000171 ***
 [3,]   0.02834     0.05220    0.543 0.587194
 [4,]   0.03384     0.04359    0.776 0.437562
 [5,]   0.06276     0.04369    1.437 0.150853
 [6,]  -0.06707     0.04147   -1.617 0.105797
 [7,]  -0.28993     0.04721   -6.141 8.20e-10 ***
 [8,]   0.41585     0.04627    8.988  < 2e-16 ***
 [9,]  -0.16783     0.05099   -3.292 0.000996 ***
[10,]   0.04792     0.04739    1.011 0.311883
[11,]  -0.08497     0.04927   -1.725 0.084564 .
[12,]  -0.07360     0.04799   -1.534 0.125097
[13,]   0.84990     0.01963   43.296  < 2e-16 ***
[14,]   0.03323     0.02406    1.381 0.167355
[15,]   0.04428     0.02444    1.812 0.070012 .
[16,]   0.82424     0.02860   28.819  < 2e-16 ***
---
Regular MA coefficient matrix
MA( 1 )-matrix
       [,1]   [,2]
```

```
[1,]  0.246 -0.168
[2,] -0.290  0.416
MA( 2 )-matrix
        [,1]   [,2]
[1,]  0.0283 0.0338
[2,] -0.1678 0.0479
MA( 3 )-matrix
        [,1]    [,2]
[1,]  0.0628 -0.0671
[2,] -0.0850 -0.0736
Seasonal MA coefficient matrix
MA( 12 )-matrix
          [,1]       [,2]
[1,] 0.8498980 0.03322685
[2,] 0.0442752 0.82423827

Residuals cov-matrix:
                    resi m1$residuals
resi         0.006941949  0.002633186
m1$residuals 0.002633186  0.007116746
----
aic=  -10.01172 ; bic=  -9.89168
> m4=refsVARMA(m3,thres=1.2)
Number of parameters:  13
initial estimates:  0.1985901 -0.08082791 0.01039118  ...
Coefficient(s):
       Estimate  Std. Error  t value Pr(>|t|)
 [1,]   0.25255     0.04521    5.586 2.32e-08 ***
 [2,]  -0.16847     0.04357   -3.866  0.00011 ***
 [3,]   0.07582     0.03910    1.939  0.05247 .
 [4,]  -0.06798     0.03953   -1.720  0.08547 .
 [5,]  -0.29222     0.04659   -6.272 3.56e-10 ***
 [6,]   0.41585     0.04668    8.909  < 2e-16 ***
 [7,]  -0.16783     0.04498   -3.732  0.00019 ***
 [8,]  -0.08497     0.04840   -1.756  0.07912 .
 [9,]  -0.07345     0.04619   -1.590  0.11182
[10,]   0.84990     0.01959   43.378  < 2e-16 ***
[11,]   0.03296     0.02395    1.376  0.16878
[12,]   0.04387     0.02441    1.797  0.07231 .
[13,]   0.82472     0.02846   28.980  < 2e-16 ***
---
Regular MA coefficient matrix
MA( 1 )-matrix
       [,1]   [,2]
[1,]  0.253 -0.168
[2,] -0.292  0.416
MA( 2 )-matrix
       [,1] [,2]
[1,]  0.000    0
[2,] -0.168    0
MA( 3 )-matrix
```

```
        [,1]    [,2]
[1,]  0.0758 -0.0680
[2,] -0.0850 -0.0734
Seasonal MA coefficient matrix
MA( 12 )-matrix
           [,1]       [,2]
[1,] 0.84989804 0.03296309
[2,] 0.04387419 0.82471809
  Residuals cov-matrix:
                    resi m1$residuals
resi         0.006964991  0.002655917
m1$residuals 0.002655917  0.007139871
----
aic=  -10.01722; bic=  -9.919685
> m5=sVARMA(zt,order=c(0,1,3),sorder=c(0,1,1),s=12,switch=T)
Regular MA coefficient matrix
MA( 1 )-matrix
       [,1]   [,2]
[1,]  0.246 -0.168
[2,] -0.297  0.416
MA( 2 )-matrix
        [,1]   [,2]
[1,]  0.0217 0.0397
[2,] -0.1678 0.0479
MA( 3 )-matrix
        [,1]    [,2]
[1,]  0.0565 -0.0615
[2,] -0.0850 -0.0675
Seasonal MA coefficient matrix
MA( 12 )-matrix
           [,1]       [,2]
[1,] 0.84989804 0.03714659
[2,] 0.02909768 0.84752096
  Residuals cov-matrix:
                    resi m1$residuals
resi         0.006938460  0.002605709
m1$residuals 0.002605709  0.007118109
----
aic=  -10.0087; bic=  -9.888657
```

## 6.2 Principal Component Analysis

Principal component analysis (PCA) is a useful tool in multivariate statistical analysis. PCA performs orthogonal rotations of the observed coordinates to seek simplification in model interpretation and/or to achieve dimension reduction in modeling. In this section we apply PCA to multivariate time series to seek stable linear relationships embedded in the data. The relationship may involve lagged variables.

To seek stable linear relations in a multivariate time series, we apply PCA to the observed series $\boldsymbol{z}_t$ and the residuals $\hat{\boldsymbol{a}}_t$ of a VAR model. The former analysis is to find stable contemporaneous

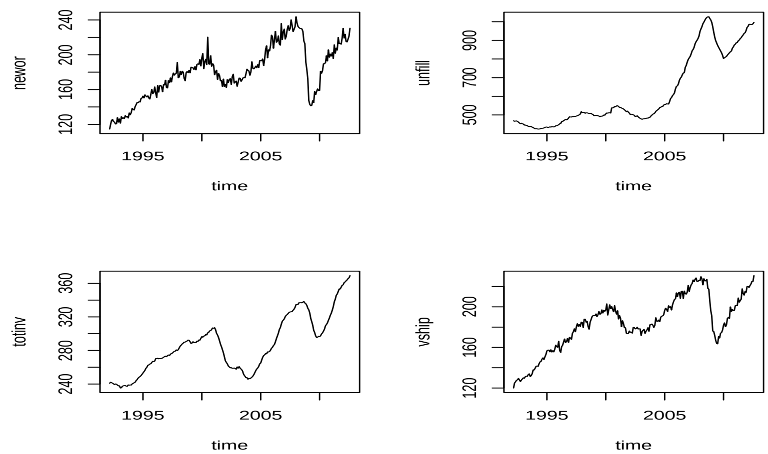


Figure 6.5: Monthly series of U.S. manufacturers data on durable goods from February 1992 to July 2012: (a) new orders, (b) total inventory, (c) unfilled orders, and (d) values of shipments. Data are in billions of dollars and seasonally adjusted.

relations where as the latter for stable lagged relations. We demonstrate the application with a real example. Consider the 4-dimensional monthly time series $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{4t})'$ of U.S. manufacturers data on durable goods, where

1. $z_{1t}$: New orders (NO),
2. $z_{2t}$: Total inventory (TI),
3. $z_{3t}$: Unfilled orders (UO),
4. $z_{4t}$: Values in shipments (VS).

All measurements are in billions of U.S. dollars and the data are seasonally adjusted. The original data are downloaded from the Federal Reserve Bank of St Louis and are in millions of dollars, and the sample period is from February 1992 to July 2012 for 246 observations. Figure 6.5 shows the time plots of the four time series. From the plots, the new orders and values of shipments show a similar trending pattern whereas total inventory and unfilled orders move closely. As expected, all four time series show an upward trend and, hence, are unit-root nonstationary.
If one specifies a VAR model for the $\boldsymbol{z}_t$ series, a VAR(4) model is selected by all order selection methods discussed in Chapter **??**. For ease in discussion, we write a VAR($p$) model as

$$\boldsymbol{z}_t = \boldsymbol{\phi}_{p,0} + \boldsymbol{\Phi}_{p,1}\boldsymbol{z}_{t-1} + \cdots + \boldsymbol{\Phi}_{p,p}\boldsymbol{z}_{t-p} + \boldsymbol{a}_{p,t}, \tag{6.6}$$

where $\boldsymbol{\phi}_{p,0}$ denotes the constant vector and the subscript "p" is used to signify the fitted order. Let $\hat{\boldsymbol{a}}_{p,t}$ be the residuals of a VAR($p$) model of Equation (6.6). Table 6.2 provides the standard deviation (square root of eigenvalue) and the proportion of variance explained by each principal component for $\boldsymbol{z}_t$ and $\hat{\boldsymbol{a}}_{p,t}$ for $p$ = 1, 2, 3, and 4. Of particular interest from the table is that

Table 6.2: Summary of Principal Component Analysis Applied to the Monthly U.S. Manufacturers Data of Durable Goods From 1992.2 to 2012.7. $\hat{\boldsymbol{a}}_{p,t}$ denotes the residuals of a VAR($p$) model.

| Series | Variable | Principal Components | | | |
|---|---|---|---|---|---|
| $\boldsymbol{z}_t$ | Stand. Dev. | 197.00 | 30.700 | 12.566 | 3.9317 |
| | Proportion | 0.9721 | 0.0236 | 0.0040 | 0.0004 |
| $\hat{\boldsymbol{a}}_{1,t}$ | Stand. Dev. | 8.8492 | 3.6874 | 1.5720 | 0.3573 |
| | Proportion | 0.8286 | 0.1439 | 0.0261 | 0.0014 |
| $\hat{\boldsymbol{a}}_{2,t}$ | Stand. Dev. | 8.3227 | 3.5233 | 1.1910 | 0.2826 |
| | Proportion | 0.8327 | 0.1492 | 0.0171 | 0.0010 |
| $\hat{\boldsymbol{a}}_{3,t}$ | Stand. Dev. | 8.0984 | 3.4506 | 1.0977 | 0.2739 |
| | Proportion | 0.8326 | 0.1512 | 0.01530 | 0.0010 |
| $\hat{\boldsymbol{a}}_{4,t}$ | Stand. Dev. | 7.8693 | 3.2794 | 1.0510 | 0.2480 |
| | Proportion | 0.8386 | 0.1456 | 0.0140 | 0.0008 |

the results of PCA appear to be stable for the residual series for all four VAR models. The last principal component explains about 0.1% of the total variability.

Table 6.3 provides the loading matrices of the principal component analyses for $\boldsymbol{z}_t$ and the residuals $\hat{\boldsymbol{a}}_{p,t}$ for $p = 1$, 2, and 3. The loading matrix for the residual series $\hat{\boldsymbol{a}}_{4,t}$ of the VAR(4) model is similar to those of the three residual series and, hence, is omitted. Again, the stability of the loading matrices for the residuals of VAR models is remarkable, especially for the last principal component. From the table, to a close approximation, the eigenvector associated with the fourth principal component can be written as $\boldsymbol{h}_4 \approx (1, 0, -1, -1)'$.
Next, consider the fitted VAR(1) model, which is

$$\boldsymbol{z}_t = \begin{bmatrix} 0.008 \\ -0.129 \\ -8.348 \\ 2.804 \end{bmatrix} + \begin{bmatrix} 0.686 & -0.027 & -0.001 & 0.357 \\ 0.116 & 0.995 & -0.000 & -0.102 \\ 0.562 & -0.023 & 0.995 & -0.441 \\ 0.108 & 0.023 & -0.003 & 0.852 \end{bmatrix} \boldsymbol{z}_{t-1} + \hat{\boldsymbol{a}}_{1,t}, \tag{6.7}$$

where all elements of the constant term $\boldsymbol{\phi}_0$ are statistically insignificant at the conventional 5% level. Pre-multiplying Equation (6.7) by $\boldsymbol{h}_4'$, we have

$$\boldsymbol{h}_4'\boldsymbol{z}_t \approx 5.55 + (0.015, -0.027, -0.994, -0.054)\boldsymbol{z}_{t-1} + \boldsymbol{h}_4'\hat{\boldsymbol{a}}_{1,t}.$$

With the small eigenvalue and the fact that means of the residuals $\hat{\boldsymbol{a}}_{1,t}$ are zero, $\boldsymbol{h}_4'\hat{\boldsymbol{a}}_{1,t}$ is essentially being zero. Consequently, with $\boldsymbol{h}_4 \approx (1, 0, -1, -1)'$, the prior equation implies

$$\mathrm{NO}_t - \mathrm{UO}_t - \mathrm{VS}_t + \mathrm{UO}_{t-1} \approx c_4,$$

where $c_4$ denotes a constant. In other words, PCA of the residuals of the VAR(1) model reveals a stable relation

$$\mathrm{NO}_t - \mathrm{VS}_t - (\mathrm{UO}_t - \mathrm{UO}_{t-1}) \approx c_4 \tag{6.8}$$

Table 6.3: Loadings of Principal Component Analysis Applied to the Monthly U.S. Manufacturers Data of Durable Goods from 1992.2 to 2012.7, where TS stands for Time Series.

| TS | Loading matrix | | | | TS | Loading matrix | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{z}_t$ | 0.102 | 0.712 | 0.342 | 0.604 | $\hat{\boldsymbol{a}}_{1,t}$ | 0.794 | 0.161 | 0.066 | 0.583 |
| | 0.152 | 0.315 | -0.928 | 0.129 | | 0.058 | -0.109 | -0.990 | 0.063 |
| | 0.978 | -0.182 | 0.098 | -0.006 | | 0.547 | -0.592 | 0.060 | -0.588 |
| | 0.096 | 0.600 | 0.110 | -0.786 | | 0.260 | 0.782 | -0.106 | -0.557 |
| $\hat{\boldsymbol{a}}_{2,t}$ | 0.796 | 0.150 | 0.017 | 0.587 | $\hat{\boldsymbol{a}}_{3,t}$ | 0.797 | 0.143 | 0.009 | 0.586 |
| | 0.026 | -0.070 | -0.997 | 0.012 | | 0.017 | -0.063 | -0.998 | 0.007 |
| | 0.543 | -0.600 | 0.049 | -0.585 | | 0.537 | -0.608 | 0.044 | -0.583 |
| | 0.267 | 0.783 | -0.055 | -0.560 | | 0.274 | 0.778 | -0.048 | -0.563 |

Therefore, for the monthly manufacturer data on durable goods, the difference between new orders and values of shipments is roughly equal to some constant plus the change in unfilled orders.

Next, consider the VAR(2) model,

$$\boldsymbol{z}_t = \widehat{\boldsymbol{\phi}}_{2,0} + \widehat{\boldsymbol{\Phi}}_{2,1}\boldsymbol{z}_{t-1} + \widehat{\boldsymbol{\Phi}}_{2,2}\boldsymbol{z}_{t-1} + \hat{\boldsymbol{a}}_{2,t}. \tag{6.9}$$

where the coefficient estimates are given in Table 6.4. Again, PCA of the residuals $\hat{\boldsymbol{a}}_{2,t}$ shows that the smallest eigenvalue is essentially zero with eigenvector approximately $\boldsymbol{h}_4 = (1, 0, -1, -1)'$. Pre-multiplying Equation (6.9) by $\boldsymbol{h}_4'$, we get

$$\boldsymbol{h}_4'\boldsymbol{z}_t \approx 2.21 + (0.59, -0.08, -1.57, -0.61)\boldsymbol{z}_{t-1} + (0.01, 0.07, 0.57, -0.01)\boldsymbol{z}_{t-2}.$$

Consequently, ignoring terms with coefficients close to zero, we have

$$z_{1t} - z_{3t} - z_{4t} - 0.59z_{1,t-1} + 1.57z_{3,t-1} + 0.61z_{4,t-1} - 0.57z_{3,t-2} \approx c_1,$$

where $c_1$ denotes a constant. Rearranging terms, the prior equation implies

$$z_{1t} - z_{3t} - z_{4t} + z_{3,t-1} - (0.59z_{1,t-1} - 0.57z_{3,t-1} - 0.61z_{4,t-1} + 0.57z_{3,t-2}) \approx c_1.$$

This approximation further simplifies as

$$(z_{1t} - z_{3t} - z_{4t} + z_{3,t-1}) - 0.59(z_{1,t-1} - z_{3,t-1} - z_{4,t-1} + z_{3,t-2}) \approx c,$$

where $c$ is a constant. The linear combination in the parentheses of the prior equation is

$$\mathrm{NO}_t - \mathrm{UO}_t - \mathrm{VS}_t + \mathrm{UO}_{t-1}$$

and its lagged value. This is exactly the linear combination in Equation (6.8). Consequently, principal component analysis of the residuals of the VAR(2) model reveals the same stable relation between the four variables of durable goods. As a matter of fact, the same stable relation is repeated when one applies PCA to the residuals of VAR(3) and VAR(4) models of $\boldsymbol{z}_t$.

***R Demonstration***: Commands used in principal component analysis

Table 6.4: Summary of the VAR(2) Model for the 4-Dimensional Time Series of Monthly Manufacturers Data on Durable Goods

| Parameter | Estimates | | | |
|---|---|---|---|---|
| $\widehat{\boldsymbol{\phi}}_{2,0}'$ | -0.221 | 3.248 | -6.267 | 3.839 |
| $\widehat{\boldsymbol{\Phi}}_{2,1}$ | 1.033 | 1.012 | -0.638 | -0.108 |
| | -0.445 | 1.549 | 0.441 | 0.537 |
| | 0.307 | 0.645 | 1.005 | -0.120 |
| | 0.141 | 0.452 | -0.072 | 0.619 |
| $\widehat{\boldsymbol{\Phi}}_{2,2}$ | 0.243 | -1.028 | 0.634 | -0.115 |
| | 0.064 | -0.568 | -0.438 | -0.166 |
| | 0.247 | -0.663 | -0.010 | -0.336 |
| | -0.016 | -0.440 | 0.070 | 0.227 |

```
> da=read.table("m-amdur.txt",header=T)
> dur= da[,3:6]/1000
> v0 =princomp(dur)
> summary(v0)
> M0 = matrix(v0$loadings[,1:4],4,4)
> VARorder(dur)  # Find VAR order
> m1=VAR(dur,1)  # Fit  VAR(1)
> v1=princomp(m1$residuals)
> summary(v1)
> M1=matrix(v1$loadings[,1:4],4,4)
> h4=matrix(c(1,0,-1,-1),4,1)
> t(h4)%*%m1$Phi
> t(h4)%*%m1$Ph0
> m2=VAR(dur,2)
> v2=princomp(m2$residuals)
> summary(v2)
> M2=matrix(v2$loadings[,1:4],4,4)
> print(round(M0,3))
> print(round(M1,3))
```

## 6.3 Use of Exogenous Variables

In many forecasting exercises, exogenous variables or independent variables are available. We discuss two approaches to handle exogenous variables in multivariate time series analysis. For simplicity, we use VAR models in the discussion.

### 6.3.1 VARX models

The first approach to include exogenous variables in multivariate time series modeling is to use the vector autoregressive model with exogenous variables. In the literature, this type of models is referred to as the VARX models with X signifying exogenous variables. The term exogenous variable is used loosely here as it may contain independent (or input) variables. Let $\boldsymbol{z}_t$ be a $k$-dimensional time series and $\boldsymbol{x}_t$ an$m$-dimensional series of exogenous variables or leading indicators. The general form of a VARX model is

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \sum_{j=0}^{s} \boldsymbol{\beta}_j \boldsymbol{x}_{t-j} + \boldsymbol{a}_t, \tag{6.10}$$

where $\boldsymbol{a}_t$ is a sequence of iid random vectors with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}_a$, $p$ and $s$ are nonnegative integers, $\boldsymbol{\phi}_i$ are the usual VAR coefficient matrices, and $\boldsymbol{\beta}_j$ are $k \times m$ coefficient matrices. This model allows for $\boldsymbol{x}_t$ to affect $\boldsymbol{z}_t$ instantaneously. The orders $p$ and $s$ of the VARX model in Equation (6.10) can be determine in several ways. For instance, one can use the information criteria or the ideal of partial F-test in multivariate multiple linear regression similar to those discussed in Chapter **??**. We use information criteria in our demonstration. The model can be estimated by the least squares method and the resulting estimates are asymptotically normally distributed. For demonstration, we consider an example.

**Example 6.2**. Consider the monthly U.S. regular conventional gas price $z_{1t}$ and No. 2 heating oil price $z_{2t}$ of New York Harbor. Both series are measured in dollars per gallon. These prices depend on the crude oil and natural gas prices. Let $x_{1t}$ be the spot oil price of West Texas Intermediate, dollars per barrel, and $x_{2t}$ the natural gas price of Henry Hub, LA, measured in dollars per million BTU. Thus, we have $\boldsymbol{z}_t = (z_{1t}, z_{2t})'$, $\boldsymbol{x}_t = (x_{1t}, x_{2t})'$, and $k = m = 2$. The sample period is from November 1993 to August 2012. We downloaded the data from the Federal Reserve Bank of St Louis. The original data of $\boldsymbol{z}_t$ are from the Energy Information Administration of U.S. Department of Energy and that of $\boldsymbol{x}_t$ are from the Wall Street Journal of Dow Jones & Company.
Figure 6.6 shows the time plots of the four monthly time series. The left panel shows the regular gas price and the heating oil price whereas the right panel contains the two independent variables. The gasoline and heating oil prices behave in a similar manner as that of the spot oil prices. The relationship between the natural gas prices and the $\boldsymbol{z}_t$ series appears to be more variable. If VAR models are entertained for the $\boldsymbol{z}_t$ series, BIC and HQ criterion selects a VAR(2) and VAR(3), respectively. The AIC, on the other hand, selects a VAR(11) model. The purpose of our demonstration is to employ a VARX model. To this end, we applied the information criteria with maximum AR order 11 and maximum lags being 3 for the input variables. Both BIC and HQ criterion select $(p, s) = (2, 1)$ for the series. If one extends the maximum lags of the input variables, then both BIC and HQ select $(p, s) = (2, 3)$, with $(p, s) =$ (2,1) as a close second. We have estimated both the VARX(2,1) and VARX(2,3) models and found that the results are similar. Therefore, we focus on the simpler VARX(2,1) model.
The selected model is then

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{\phi}_2 \boldsymbol{z}_{t-2} + \boldsymbol{\beta}_0 \boldsymbol{x}_t + \boldsymbol{\beta}_1 \boldsymbol{x}_{t-1} + \boldsymbol{a}_t. \tag{6.11}$$

Table 6.5 summarizes the estimation results of the VARX(2,1) model in Equation (6.11). The full model of the table gives unconstrained estimates whereas the reduced model shows the estimates of a simplified model. The simplification is obtained by removing simultaneously all estimates with $t$-ratio less than 1.0. Figure 6.7 shows the residual cross-correlation matrices of the residual series of the reduced model in Table 6.5. From the cross-correlation matrices, the linear dynamic dependence of the original data is largely removed. However, there are some minor residual correlations at lags

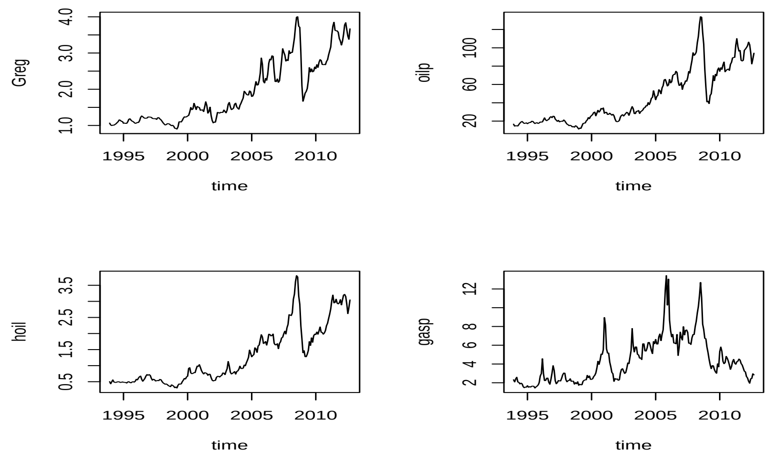


Figure 6.6: Time plots of monthly prices of energy series: The upper-left plot is the regular conventional gas price, the lower-left plot is the heating oil price, the upper-right plot is the spot oil price, and the lower-right plot is the natural gas price. The sample period is from November 1993 to August 2012.

4, 6 and 7. These residual cross- and serial correlations can also be seen from the multivariate Ljung-Box statistics of the residuals. See the plot of $p$-values of the Ljung-Box statistics in Figure 6.8. The model thus only provides a decent approximation to the underlying structure of the data.
From Table 6.5, we see that the independent variables $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$ have significant impacts on the monthly regular gas price and heating oil price. More specifically, the spot oil price and its lagged value, $x_{1t}$ and $x_{1,t-1}$, have substantial effects on the regular gas price. But the natural gas price and its lagged value, $x_{2t}$ and $x_{2,t-1}$ only have marginal effects on the regular gas price. On the other hand, both $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$ affect significantly the heating oil price. The importance of the independent variables $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$ can also be seen by comparing the VARX(2,1) model in Equation (6.11) with a pure VAR(2) model. The residual covariance matrix of a VAR(2) model for $\boldsymbol{z}_t$ is

$$\widehat{\boldsymbol{\Sigma}}_{\text{var}} = 10^2 \begin{bmatrix} 1.327 & 0.987 \\ 0.987 & 1.463 \end{bmatrix}.$$

Clearly, the residual variances of the VAR(2) model are much larger than those of the VARX(2,1) model.

**Remark**: Analysis of VARX models is carried out by the commands `VARXorder`, `VARX`, and `refVARX` of the `MTS` package. `VARXorder` provides information criteria for data, `VARX` performs least squares estimation of a specified VARX$(p, s)$ model, and `refVARX` refines a fitted VARX model by removing insignificant estimates.

***R Demonstration***: VARX modeling. Output edited.

Table 6.5: Estimation Results of a VARX(2,1) Model for the Monthly Gas and Heating Oil Prices with Spot Oil and Natural Gas Prices as Input Variables. The Sample Period is from November 1993 to August 2012.

| | Full Model | | | | Reduced Model | | | |
|---|---|---|---|---|---|---|---|---|
| Par. | Estimates | | St.Errors | | Estimates | | St.Errors | |
| $\boldsymbol{\phi}_0'$ | 0.182 | -0.017 | 0.030 | 0.019 | 0.183 | 0 | 0.028 | 0 |
| $\boldsymbol{\phi}_1$ | 1.041 | 0.169 | 0.059 | 0.080 | 1.044 | 0.162 | 0.054 | 0.056 |
| | 0.005 | 0.844 | 0.037 | 0.050 | 0 | 0.873 | 0 | 0.031 |
| $\boldsymbol{\phi}_2$ | -0.322 | -0.008 | 0.056 | 0.068 | -0.327 | 0 | 0.040 | 0 |
| | 0.014 | 0.012 | 0.035 | 0.043 | 0 | 0 | 0 | 0 |
| $\boldsymbol{\beta}_0$ | 0.018 | 0.010 | 0.001 | 0.007 | 0.018 | 0.010 | 0.001 | 0.007 |
| | 0.024 | 0.026 | 0.001 | 0.004 | 0.023 | 0.026 | 0.001 | 0.004 |
| $\boldsymbol{\beta}_1$ | -0.015 | -0.012 | 0.002 | 0.007 | -0.015 | -0.012 | 0.002 | 0.007 |
| | -0.019 | -0.029 | 0.001 | 0.004 | -0.019 | -0.029 | 0.001 | 0.004 |
| $10^2\boldsymbol{\Sigma}_a$ | 0.674 | 0.105 | | | 0.674 | 0.105 | | |
| | 0.105 | 0.268 | | | 0.105 | 0.270 | | |

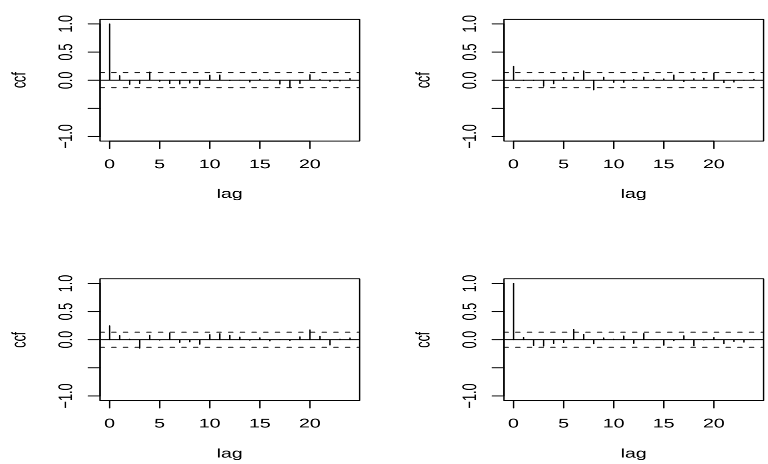


Figure 6.7: Residual cross-correlation matrices of the reduced VARX(2,1) model in Table 6.5.

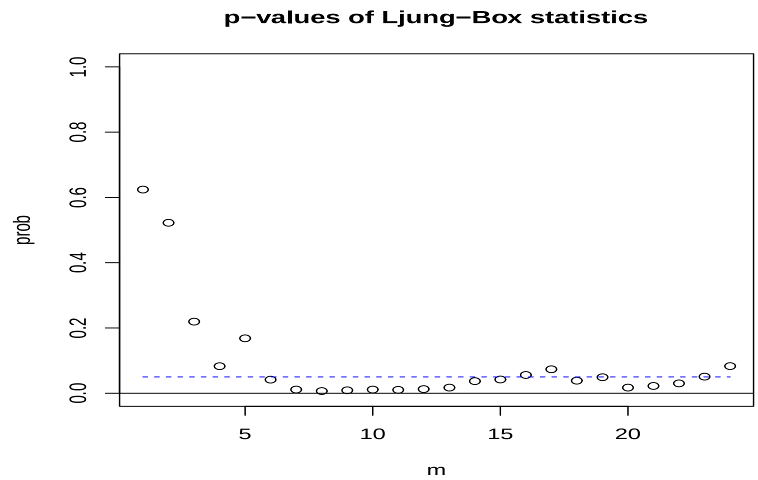


Figure 6.8: $p$-values of multivariate Ljung-Box statistics of the residual series of the reduced VARX(2,1) model in Table 6.5.

```
> da=read.table("m-gasoil.txt",header=T)
> head(da)
  year mon  Greg  hoil   oilp gasp
1 1993  11 1.066 0.502 16.699 2.32
 ....
6 1994   4 1.027 0.479 16.380 2.04
> zt=da[,3:4]; xt=da[,5:6]
> VARorder(zt)
selected order: aic =  11
selected order: bic =  2
selected order: hq =  3
> VARXorder(zt,exgo,maxp=11,maxm=2) # order selection
selected order(p,s): aic =  11 1
selected order(p,s): bic =  2 1
selected order(p,s): hq =  2 1
> m1=VARX(zt,2,xt,1) # Full model estimation
constant term:
est:  0.1817 -0.017
 se:  0.0295 0.0186
AR( 1 ) matrix
      Greg  hoil
Greg 1.041 0.169
hoil 0.005 0.844
standard errors
      [,1] [,2]
[1,] 0.059 0.08
[2,] 0.037 0.05
AR( 2 ) matrix
       Greg   hoil
```

```
Greg -0.322 -0.008
hoil  0.014  0.012
standard errors
      [,1]  [,2]
[1,] 0.056 0.068
[2,] 0.035 0.043
Coefficients of exogenous
lag- 0  coefficient matrix
      oilp  gasp
Greg 0.018 0.010
hoil 0.024 0.026
standard errors
      [,1]  [,2]
[1,] 0.001 0.007
[2,] 0.001 0.004
lag- 1  coefficient matrix
       oilp   gasp
Greg -0.015 -0.012
hoil -0.019 -0.029
standard errors
      [,1]  [,2]
[1,] 0.002 0.007
[2,] 0.001 0.004
Information criteria:
AIC:  -10.83396; BIC:  -10.55981
> cov(m2$residuals)
            Greg        hoil
Greg 0.006736072 0.001049648
hoil 0.001049648 0.002678639
> m1a=refVARX(m1,thres=1) # Model refinement
constant term:
est:  0.1828 0
 se:  0.028 1
AR( 1 ) matrix
      [,1]  [,2]
[1,] 1.044 0.162
[2,] 0.000 0.873
standard errors
      [,1]  [,2]
[1,] 0.054 0.056
[2,] 1.000 0.031
AR( 2 ) matrix
       [,1] [,2]
[1,] -0.327    0
[2,]  0.000    0
standard errors
     [,1] [,2]
[1,] 0.04    1
[2,] 1.00    1
Coefficients of exogenous
lag- 0  coefficient matrix
```

```
      [,1]  [,2]
[1,] 0.018 0.010
[2,] 0.023 0.026
standard errors
      [,1]  [,2]
[1,] 0.001 0.007
[2,] 0.001 0.004
lag- 1  coefficient matrix
       [,1]   [,2]
[1,] -0.015 -0.012
[2,] -0.019 -0.029
standard errors
      [,1]  [,2]
[1,] 0.002 0.007
[2,] 0.001 0.004
Information criteria:
AIC:  -10.87015
BIC:  -10.67216
> MTSdiag(m1a)  # Model checking
[1] "Covariance matrix:"
        Greg    hoil
Greg 0.00674 0.00105
hoil 0.00105 0.00270
```

### 6.3.2 Regression Model

The second approach to include independent variables in multivariate time series analysis is to use multivariate linear regression models with time series errors. The model can be written as

$$\boldsymbol{\phi}(B)(\boldsymbol{z}_t - \boldsymbol{\beta}\boldsymbol{w}_t) = \boldsymbol{a}_t, \tag{6.12}$$

where $\boldsymbol{a}_t$ is defined in Equation (6.10), $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$ is a matrix polynomial with nonnegative degree $p$, $\boldsymbol{w}_t$ is a $v$-dimensional vector of independent variables, and $\boldsymbol{\beta}$ is a $k \times v$ coefficient matrix. In most applications, we have $w_{1t} = 1$ for all $t$ so that a constant term is included in the model. Here $\boldsymbol{w}_t$ may contain lagged values of the observed independent variable $\boldsymbol{x}_t$. The model in Equation (6.12) is nonlinear in parameters and differs from the VARX model of Equation (6.10) in several ways, even if $\boldsymbol{w}_t = (1, \boldsymbol{x}_t')'$. To see this, assume that $\boldsymbol{w}_t = (1, \boldsymbol{x}_t')'$ so that $v = m + 1$. Partition the coefficient matrix $\boldsymbol{\beta}$ as $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$ with $\boldsymbol{\beta}_1$ being a column vector. Then, we can rewrite the model in Equation (6.12) as

$$\begin{aligned} \boldsymbol{z}_t &= \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \boldsymbol{\phi}(1)\boldsymbol{\beta}_1 + \boldsymbol{\phi}(B)\boldsymbol{\beta}_2 \boldsymbol{x}_t + \boldsymbol{a}_t \\ &\equiv \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \sum_{j=0}^{p} \boldsymbol{\gamma}_j \boldsymbol{x}_{t-j} + \boldsymbol{a}_t, \end{aligned} \tag{6.13}$$

where $\boldsymbol{\phi}_0 = \boldsymbol{\phi}(1)\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_2$ and $\boldsymbol{\gamma}_j = -\boldsymbol{\phi}_j \boldsymbol{\beta}_2$ for $j = 1, \ldots, p$. From Equation (6.13), $\boldsymbol{z}_t$ also depends on the lagged values of $\boldsymbol{x}_t$ provided $p > 0$. Thus, the regression model with time series errors of Equation (6.13) can be regarded as a special case of the VARX model in Equation (6.10).

On the other hand, due to its parameter constraints, the model in Equation (6.12) requires nonlinear estimation. The likelihood method is commonly used.
In applications, a two-step procedure is often used to specify a regression model with time series errors. In the first step, the multivariate multiple linear regression

$$\boldsymbol{z}_t = \boldsymbol{c} + \boldsymbol{\beta}\boldsymbol{x}_t + \boldsymbol{e}_t,$$

is fitted by the ordinary least squares method to obtain the residuals $\hat{\boldsymbol{e}}_t$. In the second step, information criteria are used to select a VAR model for $\hat{\boldsymbol{e}}_t$.

**Example 6.2 (continued)**. To demonstrate, we again consider the monthly prices of regular gas and heating oil with spot price of crude oil and price of natural gas as independent variables. We start with fitting the multivariate linear regression model,

$$\boldsymbol{z}_t = \begin{bmatrix} 0.627 \\ -0.027 \end{bmatrix} + \begin{bmatrix} 0.029 & -0.007 \\ 0.030 & -0.008 \end{bmatrix} \boldsymbol{x}_t + \hat{\boldsymbol{e}}_t,$$

and use the residual series $\hat{\boldsymbol{e}}_t$ to specify the VAR order. All three information criteria employed identify a VAR(2) model. Thus, we use the model

$$(\boldsymbol{I}_2 - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2)(\boldsymbol{z}_t - \boldsymbol{c} - \boldsymbol{\beta}\boldsymbol{x}_t) = \boldsymbol{a}_t, \tag{6.14}$$

where $\boldsymbol{c}$ is a 2-dimensional constant vector. Table 6.6 summarizes the estimation results of the model in Equation (6.14). The full model denotes the maximum likelihood estimates with bivariate Gaussian distribution and the reduced model is obtained by setting simultaneously all estimates with $t$-ratio less than 1.2 to zero. The choice of the threshold 1.2 is based on the AIC as several threshold values were used. Figure 6.9 shows the residual cross-correlation matrices of the reduced model in Table 6.6 whereas Figure 6.10 plots the $p$-values of the multivariate Ljung-Box statistics for the residuals of the reduced model. Both figures indicate that the reduced model in Table 6.6 is adequate in describing the linear dynamic dependence of the two price series.

It is interesting to compare the reduced VARX(2,1) model of Table 6.5 with that of Table 6.6. First, the VARX model employs 13 coefficients whereas the regression model only uses 9 coefficients. The regression model is more parsimonious. However, the information criteria seem to prefer the VARX(2,1) model; see the R demonstrations. Second, the two reduced models are not nested because the regression model implicitly employs $\boldsymbol{x}_{t-2}$. In the form of Equation (6.13), the reduced model of Table 6.6 approximately becomes

$$\begin{aligned} \boldsymbol{z}_t &= \begin{bmatrix} 0.201 \\ -0.030 \end{bmatrix} + \begin{bmatrix} 1.009 & 0 \\ 0.045 & 0.973 \end{bmatrix} \boldsymbol{z}_{t-1} + \begin{bmatrix} -0.305 & 0.166 \\ 0 & -0.128 \end{bmatrix} \boldsymbol{z}_{t-2} \\ &= \begin{bmatrix} 0.027 & 0 \\ 0.028 & 0 \end{bmatrix} \boldsymbol{x}_t - \begin{bmatrix} 0.027 & 0 \\ 0.028 & 0 \end{bmatrix} \boldsymbol{x}_{t-1} + \begin{bmatrix} 0.003 & 0 \\ 0.004 & 0 \end{bmatrix} \boldsymbol{x}_{t-2} + \boldsymbol{a}_t. \end{aligned} \tag{6.15}$$

From the prior model representation, the added impact of $\boldsymbol{x}_{t-2}$ on $\boldsymbol{z}_t$ appears to be small. Consequently, the regression model is approximately a sub-model of the VARX(2,1) model. It is then not surprising to see that the residual variances of the VARX(2,1) model are smaller than those of the regression model. Third, the model in Equation (6.15) seems to suggest that, conditional on the spot price of crude oil and lagged values of $\boldsymbol{z}_t$, the prices of regular gas and heating oil do

Table 6.6: Estimation Results of the Regression Model with Time Series Errors in Equation (6.14) for the Monthly Prices of Regular Gas and Heating Oil with Spot Price of Crude Oil and Price of Natural Gas as Independent Variable.

| | Full Model | | | | Reduced Model | | | |
|---|---|---|---|---|---|---|---|---|
| Par. | Estimate | | St.Errors | | Estimate | | St.Errors | |
| $\boldsymbol{c}'$ | 0.650 | -0.025 | 0.049 | 0.043 | 0.679 | 0 | 0.036 | 0 |
| $\boldsymbol{\beta}$ | 0.028 | -0.005 | 0.001 | 0.008 | 0.027 | 0 | 0.001 | 0 |
| | 0.029 | -0.004 | 0.001 | 0.006 | 0.028 | 0 | 0.001 | 0 |
| $\boldsymbol{\phi}_1$ | 1.012 | -0.071 | 0.069 | 0.113 | 1.009 | 0 | 0.068 | 0 |
| | 0.070 | 0.949 | 0.044 | 0.071 | 0.045 | 0.973 | 0.030 | 0.066 |
| $\boldsymbol{\phi}_2$ | -0.326 | 0.186 | 0.068 | 0.110 | -0.305 | 0.166 | 0.064 | 0.070 |
| | -0.046 | -0.138 | 0.044 | 0.072 | 0 | -0.128 | 0 | 0.073 |
| $10^2\boldsymbol{\Sigma}_a$ | 0.92 | 0.19 | | | 0.897 | 0.171 | | |
| | 0.19 | 0.38 | | | 0.171 | 0.355 | | |

not depend on the price of natural gas. The VARX(2,1) model does not reveal such a structure. However, Equation (6.15) indicates that there exists a feedback relation between prices of regular gas and heating oil. The reduced VARX model, on the other hand, shows that, conditional of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$, the price of natural gas does not depend on the past values of the regular gas price.

**Remark**: Multivariate regression models with time series errors can be estimated by the commands `REGts` and `refREGts` of the `MTS` package. Multivariate multiple linear regression can be estimated by the multivariate linear model command `Mlm`.

***R Demonstration***: Regression model with time series errors. Output edited.

```
> da=read.table("m-gasoil.txt",header=T)
> zt=da[,3:4]; xt=da[,5:6]
> m1=Mlm(zt,xt)
[1] "LSE of parameters"
[1] "  est   s.d.   t-ratio   prob"
         [,1]     [,2]  [,3]   [,4]
[1,]  0.62686 0.022768 27.53 0.0000
[2,]  0.02864 0.000424 67.59 0.0000
[3,] -0.00675 0.005222 -1.29 0.1977
[4,] -0.02659 0.016308 -1.63 0.1044
[5,]  0.02961 0.000303 97.56 0.0000
[6,] -0.00846 0.003740 -2.26 0.0246
> VARorder(m1$residuals)  # Order selection
selected order: aic =  2
selected order: bic =  2
selected order: hq =  2
> m3=REGts(zt,2,xt) # Estimation
Number of parameters:  14
```

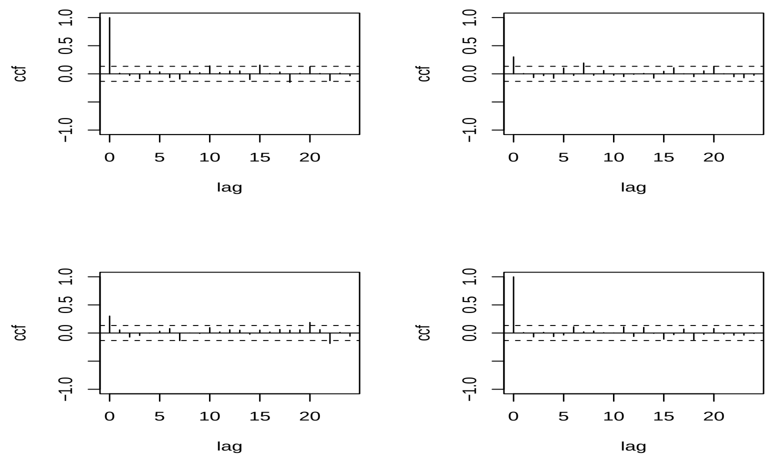


Figure 6.9: Residual cross-correlation matrices of the reduced regression model in Table 6.6.

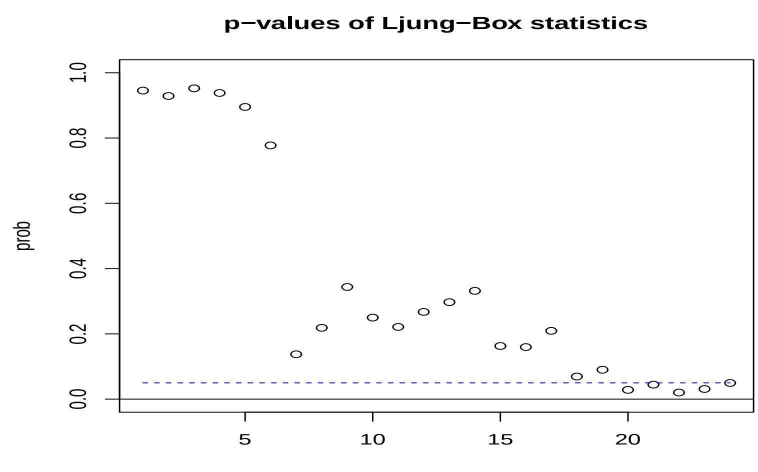


Figure 6.10: $p$-values of multivariate Ljung-Box statistics of the residual series of the reduced regression model in Table 6.6.

```
=======
Coefficient matrix for constant + exogenous variable
Estimates:
       [,1]  [,2]   [,3]
[1,]  0.650 0.028 -0.005
[2,] -0.025 0.029 -0.004
Standard errors:
      [,1]  [,2]  [,3]
[1,] 0.049 0.001 0.008
[2,] 0.043 0.001 0.006
VAR coefficient matrices:
AR( 1 ) coefficient:
      [,1]   [,2]
[1,] 1.012 -0.071
[2,] 0.070  0.949
standard errors:
      [,1]  [,2]
[1,] 0.069 0.113
[2,] 0.044 0.071
AR( 2 ) coefficient:
       [,1]   [,2]
[1,] -0.326  0.186
[2,] -0.046 -0.138
standard errors:
      [,1]  [,2]
[1,] 0.068 0.110
[2,] 0.044 0.072
Residual Covariance matrix:
       Greg   hoil
Greg 0.0092 0.0019
hoil 0.0019 0.0038
============
Information criteria:
AIC:  -10.26927; BIC:  -10.05605
> m3c=refREGts(m3,thres=1.2)  # Model refinement
Number of parameters:  9
=======
Coefficient matrix for constant + exogenous variable
Estimates:
      [,1]  [,2] [,3]
[1,] 0.679 0.027    0
[2,] 0.000 0.028    0
Standard errors:
      [,1]  [,2] [,3]
[1,] 0.036 0.001    1
[2,] 1.000 0.001    1
VAR coefficient matrices:
AR( 1 ) coefficient:
      [,1]  [,2]
[1,] 1.009 0.000
[2,] 0.045 0.973
```

```
standard errors:
      [,1]  [,2]
[1,] 0.068 1.000
[2,] 0.030 0.066
AR( 2 ) coefficient:
       [,1]   [,2]
[1,] -0.305  0.166
[2,]  0.000 -0.128
standard errors:
      [,1]  [,2]
[1,] 0.064 0.070
[2,] 1.000 0.073
Residual Covariance matrix:
       Greg   hoil
Greg 0.0089 0.0017
hoil 0.0017 0.0035
============
Information criteria:
AIC:  -10.37776; BIC:  -10.24069
> MTSdiag(m3c)  # Model checking
```

## 6.4 Factor Models

We briefly discuss some factor models.

### 6.4.1 Orthogonal Factor Models

The simplest factor model is the traditional orthogonal factor model in which the observed time series $\boldsymbol{z}_t$ is driven by a small number of common factors. The model can be written as

$$\boldsymbol{z}_t = \boldsymbol{L}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{6.16}$$

where $\boldsymbol{L}$ is a $k \times m$ loading matrix, $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{mt})'$ with $m << k$ is an $m$-dimensional vector of common factors, and $\boldsymbol{\epsilon}_t$ is a sequence of $k$-dimensional random vectors with mean zero and Cov($\boldsymbol{\epsilon}_t$) = $\boldsymbol{\Sigma}_\epsilon$ = diag$\{\sigma_1^2, \ldots, \sigma_k^2\}$ with "diag" denoting diagnoal matrix. In Equation (6.16), $\boldsymbol{f}_t$ and $\boldsymbol{\epsilon}_t$ are assumed to be independent and, for identifiability, we require that $E(\boldsymbol{f}_t) = \boldsymbol{0}$ and Cov($\boldsymbol{f}_t$) = $\boldsymbol{I}_{m\times m}$. The loading matrix $\boldsymbol{L}$ is assumed to be of full column rank; otherwise, the number of common factors can be reduced. In Equation (6.16), we assume, for simplicity, $E(\boldsymbol{z}_t) = \boldsymbol{0}$.
Under the aforementioned assumptions and stationarity, the autocovariance matrices of $\boldsymbol{z}_t$ satisfy

$$\begin{aligned} \boldsymbol{\Gamma}_z(0) &= \boldsymbol{L}\boldsymbol{\Gamma}_f(0)\boldsymbol{L}' + \boldsymbol{\Sigma}_\epsilon = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Sigma}_\epsilon \\ \boldsymbol{\Gamma}_z(\ell) &= \boldsymbol{L}\boldsymbol{\Gamma}_f(\ell)\boldsymbol{L}', \quad \ell > 0, \end{aligned} \tag{6.17}$$

where $\boldsymbol{\Gamma}_y(\ell)$ denotes the lag-$\ell$ autocovariance matrix of a vector time series $\boldsymbol{y}_t$. From Equation (6.17), the autocovariance matrices of an orthogonal factor model $\boldsymbol{z}_t$ are all singular for $\ell > 0$. This leads to the idea of defining the number of common factor $m$ as the maximum of the ranks of $\boldsymbol{\Gamma}_z(\ell)$ for $\ell > 0$. Statistical methods for exploiting the property in Equation (6.17) were proposed by Peña and Box (1987) and Geweke (1977). A weakness of this approach is that Equation (6.17) holds only for stationary processes $\boldsymbol{z}_t$. Peña and Poncela (2006) extended the model to include unit-root nonstationary processes.

Estimation of the orthogonal factor models in Equation (6.16) can be carried out in two ways. The first approach is to apply principal component (PC) analysis to $\boldsymbol{z}_t$ and select the first $m$ PCs corresponding to the largest eigenvalues as the common factors. From the spectral analysis, we have

$$\boldsymbol{\Gamma}_z(0) = \lambda_1^2 \boldsymbol{e}_1 \boldsymbol{e}_1' + \lambda_2^2 \boldsymbol{e}_2 \boldsymbol{e}_2' + \cdots + \lambda_k^2 \boldsymbol{e}_k \boldsymbol{e}_k',$$

where $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_k^2$ are the eigenvalues of the covariance matrix $\boldsymbol{\Gamma}_z(0)$ and $\boldsymbol{e}_i$ is an eigenvector associated with the eigenvalue $\lambda_i^2$ and satisfies $\|\boldsymbol{e}_i\|^2 = 1$. Suppose that the first $m$ eigenvalues are large and the other eigenvalues are small, then we can obtain the approximation

$$\boldsymbol{\Gamma}_z(0) \approx [\lambda_1 \boldsymbol{e}_1, \lambda_2 \boldsymbol{e}_2, \ldots, \lambda_m \boldsymbol{e}_m] \begin{bmatrix} \lambda_1 \boldsymbol{e}_1' \\ \lambda_2 \boldsymbol{e}_2' \\ \vdots \\ \lambda_m \boldsymbol{e}_m' \end{bmatrix} + \widehat{\boldsymbol{\Sigma}}_\epsilon \equiv \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}' + \widehat{\boldsymbol{\Sigma}}_\epsilon, \tag{6.18}$$

where $\widehat{\boldsymbol{\Sigma}}_e = \text{diag}\{u_1^2, \ldots, u_k^2\} = \text{diag}[\boldsymbol{\Gamma}_z(0)] - \text{diag}[\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}']$. Traditionally, the choice of $m$ is determined subjectively by looking at the scree plot of eigenvalues of $\boldsymbol{\Gamma}_z(0)$. To avoid the scaling effect, one may choose to standardize the components of $\boldsymbol{z}_t$ before estimating the factor model. This is equivalent to replacing $\boldsymbol{\Gamma}_z(0)$ by the correlation matrix of $\boldsymbol{z}_t$. For further details of using PCA in factor analysis, see Johnson and Wichern (2007), Peña and Box (1987), and the diffusion index discussed below. The second approach is to use maximum likelihood estimation under the normality assumption and some regularity conditions for model identification. This approach also requires that $m$ is known. Commonly used regularity conditions for maximum likelihood estimation are as follows: (a) $\boldsymbol{\Sigma}_\epsilon$ is invertiable and (b) $\boldsymbol{L}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{L}$ is a diagonal matrix.

**Example 6.4**. Consider the monthly log returns of stocks for ten U.S. companies from January 2001 to December 2011 for 132 observations. The companies and their return summary statistics are given in Table 6.7. These ten companies can roughly be classified into three industrial sectors, namely semiconductor, pharmaceutical, and investment banks. The mean returns of the ten stocks are all close to zero, but the log returns have some negative skewness and high excess kurtosis.

Let $\boldsymbol{z}_t$ be the standardized log returns, where the $i$th component is defined as $z_{it} = r_{it}/\hat{\sigma}_i$ with $r_{it}$ being the $i$th return series and $\hat{\sigma}_i$ being the sample standard error of $r_{it}$. We perform the principal component analysis of $\boldsymbol{z}_t$. Figure 6.11 shows the scree plot of the analysis. From the plot, eigenvalues of the sample correlation matrix of $\boldsymbol{z}_t$ decay exponentially and the first few eigenvalues appear to be substantially larger than the others. Thus, one may select $m$ = 3 or 4. We use $m$ = 3 for simplicity. The PCA shows that the first three components explain about 72.4% of the total variability of the log returns. See the attached R output for details. Estimates of the loading matrix $\boldsymbol{L}$ and the covariance matrix of $\boldsymbol{\epsilon}_t$ are given in Table 6.8.

From Equation (6.18), the $i$th column of the loading matrix $\boldsymbol{L}$ is proportional to the $i$th normalized eigenvector which, in turn, produces the $i$th principal component of the data. The estimated loading matrix $\boldsymbol{L}$ in Table 6.8 suggests that (a) the first factor is an weighted average of the log returns with similar weights for stocks in the same industrial sector, (b) the second factor is essentially a weighted difference of the log returns between the semiconductor sector and the pharmaceutical industry, and (c) the third factor represents a weighted difference of log returns between the semiconductor

Table 6.7: Ten U.S. Companies and Some Summary Statistics of Their Monthly Log Returns from January 2001 to December 2011.

| Sector | Name | Tick | Mean | Var. | Skew. | Kurt. |
|---|---|---|---|---|---|---|
| Semi- | Texas Instru. | TXN | -0.003 | 0.012 | -0.523 | 0.948 |
| cond. | Micron Tech. | MU | -0.013 | 0.027 | -0.793 | 2.007 |
| | Intel Corp. | INTC | -0.000 | 0.012 | -0.584 | 1.440 |
| | Taiwan Semi. | TSM | 0.006 | 0.013 | -0.201 | 1.091 |
| Pharma- | Pfizer | PFE | -0.003 | 0.004 | -0.284 | -0.110 |
| ceutical | Merck & Co. | MRK | -0.003 | 0.006 | -0.471 | 1.017 |
| | Eli Lilly | LLY | -0.003 | 0.005 | -0.224 | 1.911 |
| Invest. | JPMorgan Chase | JPM | 0.000 | 0.009 | -0.466 | 1.258 |
| Bank | Morgan Stanley | MS | -0.010 | 0.013 | -1.072 | 4.027 |
| | Goldman Sachs | GS | -0.001 | 0.008 | -0.500 | 0.455 |

sector and the investment banks. On the other hand, the variances of the noise components are between 18 to 35% of each standardized return series, indicating that marked variability remains in each log return series. Figure 6.12 shows the time plots of the three common factors, i.e. $\hat{\boldsymbol{f}}_t$. We also applied the maximum likelihood estimation of orthogonal factor models to the standardized log returns, assuming $m = 3$. The results are also given in Table 6.8. No factor rotation is used so that one can easily compares the results with those of the principal component approach. Indeed, the estimation results are similar, at least qualitatively. The two estimated loading matrices are close. The main difference appears to be in the variances of the innovations. The maximum likelihood method shows that the variances of individual innovations for MU, PFE, MRK, and JPM are around 50%. This indicates the choice of 3 common factor might be too low. As a matter of fact, a choice of $m = 5$ appears to provide a better fit in this particular instance.

***R Demonstration***: Example of orthogonal factor model.

```
> da=read.table("m-tenstocks.txt",header=T)
> rtn=log(da[,2:11]+1)  # log returns
> std=diag(1/sqrt(diag(cov(rtn))))
> rtns=as.matrix(rtn)%*%std   # Standardize individual series
> m1=princomp(rtns)
> names(m1)
[1] "sdev"   "loadings" "center"  "scale"   "n.obs"    "scores"   "call"
> sdev=m1$sdev    # square root of eigenvalues
> M=m1$loadings
> summary(m1)
Importance of components:
                        Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
Standard deviation     2.14257 1.292276 0.9617832 0.8581358 0.7543204
Proportion of Variance 0.46257 0.168273 0.0932088 0.0742018 0.0573343
Cumulative Proportion  0.46257 0.630839 0.7240474 0.7982492 0.8555835
                        Comp.6     Comp.7    Comp.8    Comp.9   Comp.10
Standard deviation     0.6351810 0.5690478 0.5351137 0.4738664 0.441654
```

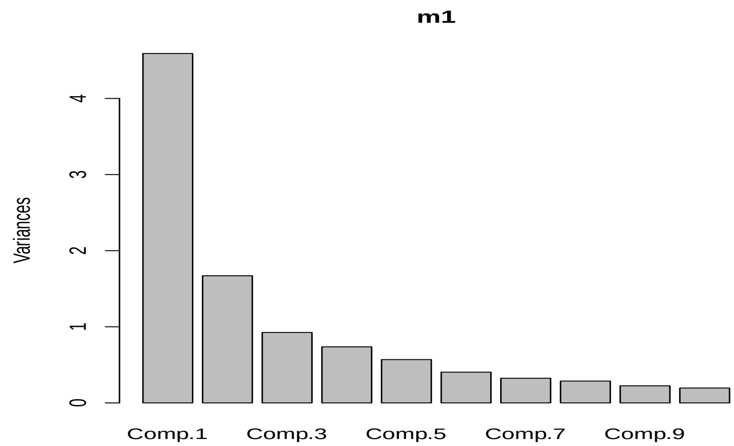


Figure 6.11: Scree plot for the monthly log returns of ten U.S. companies given in Table 6.7.

Table 6.8: Estimation Results of Orthogonal Factor Models for the Monthly Log Returns of Ten U.S. Companies given in Table 6.7, where $\boldsymbol{L}_i$ is the $i$th column of $\boldsymbol{L}$ and $\Sigma_{\epsilon,i}$ is the $(i,i)$th element of $\boldsymbol{\Sigma}_\epsilon$.

| | PCA Approach | | | | Maximum Likelihood Method | | | |
|---|---|---|---|---|---|---|---|---|
| Tick | $\boldsymbol{L}_1$ | $\boldsymbol{L}_2$ | $\boldsymbol{L}_3$ | $\Sigma_{\epsilon,i}$ | $\boldsymbol{L}_1$ | $\boldsymbol{L}_2$ | $\boldsymbol{L}_3$ | $\Sigma_{\epsilon,i}$ |
| TXN | 0.788 | 0.198 | 0.320 | 0.237 | 0.714 | -0.132 | 0.417 | 0.299 |
| MU | 0.671 | 0.361 | 0.289 | 0.336 | 0.579 | -0.263 | 0.315 | 0.496 |
| INTC | 0.789 | 0.183 | 0.333 | 0.232 | 0.718 | -0.130 | 0.442 | 0.272 |
| TSM | 0.802 | 0.270 | 0.159 | 0.258 | 0.725 | -0.186 | 0.330 | 0.331 |
| PFE | 0.491 | -0.643 | -0.027 | 0.345 | 0.351 | 0.580 | 0.197 | 0.501 |
| MRK | 0.402 | -0.689 | 0.226 | 0.312 | 0.282 | 0.582 | 0.220 | 0.533 |
| LLY | 0.448 | -0.698 | 0.061 | 0.309 | 0.359 | 0.638 | 0.184 | 0.430 |
| JPM | 0.724 | 0.020 | -0.345 | 0.357 | 0.670 | 0.067 | 0.010 | 0.547 |
| MS | 0.755 | 0.053 | -0.425 | 0.246 | 0.789 | 0.052 | -0.154 | 0.352 |
| GS | 0.745 | 0.122 | -0.498 | 0.182 | 0.879 | -0.008 | -0.363 | 0.096 |

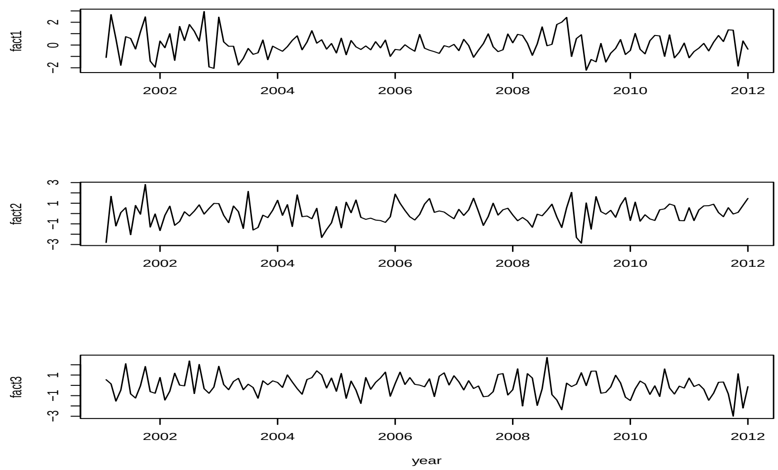


Figure 6.12: Estimated common factors for the monthly log returns of ten U.S. companies given in Table 6.7.

```
Proportion of Variance 0.0406535 0.0326287 0.0288533 0.0226263 0.019655
Cumulative Proportion  0.8962370 0.9288657 0.9577190 0.9803453 1.000000
> screeplot(m1)
> SD=diag(sdev[1:3])  # Compute the loading matrix
> L=M[,1:3]%*%SD
> print(round(L,3))
        [,1]   [,2]   [,3]
 [1,] -0.788 -0.198 -0.320
 [2,] -0.671 -0.361 -0.289
 [3,] -0.789 -0.183 -0.333
 [4,] -0.802 -0.270 -0.159
 [5,] -0.491  0.643  0.027
 [6,] -0.402  0.689 -0.226
 [7,] -0.448  0.698 -0.061
 [8,] -0.724 -0.020  0.345
 [9,] -0.755 -0.053  0.425
[10,] -0.745 -0.122  0.498
> LLt=L%*%t(L)
> diag(LLt)
 [1] 0.762621 0.664152 0.767795 0.741557 0.655427 0.687690 0.691248
 [8] 0.643403 0.753815 0.817913
> SigE=1-diag(LLt)
> SigE
 [1] 0.237379 0.335848 0.232205 0.258444 0.344573 0.312310 0.308752
 [8] 0.356597 0.246185 0.182087

> m2=factanal(rtns,3,scores="regression",rotation="none") #MLE
> m2
factanal(x =rtns, factors=3, scores ="regression", rotation ="none")
```

```
Uniquenesses:
 [1] 0.299 0.496 0.272 0.331 0.501 0.533 0.430 0.547 0.352 0.096
Loadings:
      Factor1 Factor2 Factor3
 [1,]  0.714  -0.132   0.417
 [2,]  0.579  -0.263   0.315
....
> names(m2)
 [1] "converged" "loadings"  "uniquenesses" "correlation" "criteria"
 [6] "factors"   "dof"       "method"       "scores"      "STATISTIC"
[11] "PVAL"      "n.obs"     "call"
> L2=matrix(m2$loadings,10,3)
> print(round(L2,3))
       [,1]   [,2]   [,3]
 [1,] 0.714 -0.132  0.417
 [2,] 0.579 -0.263  0.315
 [3,] 0.718 -0.130  0.442
 [4,] 0.725 -0.186  0.330
 [5,] 0.351  0.580  0.197
 [6,] 0.282  0.582  0.220
 [7,] 0.359  0.638  0.184
 [8,] 0.670  0.067  0.010
 [9,] 0.789  0.052 -0.154
[10,] 0.879 -0.008 -0.363
```

**Discussion**: The PCA approach to estimating orthogonal factor models in Equation (6.18) provides only an approximate solution in general. However, the approach can be justified asymptotically if the covariance matrix of the noises is proportional to the $k \times k$ identity matrix, i.e., $\boldsymbol{\Sigma}_\epsilon = \sigma^2 \boldsymbol{I}_k$. To see this, we consider the simplest case of $k = 2$, $m = 1$, and $E(\boldsymbol{z}_t) = \boldsymbol{0}$. In this case, the model is

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} f_t + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \tag{6.19}$$

where $\text{Cov}(\boldsymbol{\epsilon}_t) = \sigma^2 \boldsymbol{I}_2$, $L_1$ and $L_2$ are real numbers and $f_t$ is the common factor. Suppose the sample size is $T$. Then, we have

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \frac{1}{T}\left(\sum_{t=1}^{T} f_t^2\right)[L_1, L_2] + \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'. \tag{6.20}$$

Under the assumptions of orthogonal factor models, Equation (6.20) is asymptotically equivalent to

$$\boldsymbol{\Gamma}_z(0) = \begin{bmatrix} L_1^2 & L_1 L_2 \\ L_1 L_2 & L_2^2 \end{bmatrix} + \sigma^2 \boldsymbol{I}_2.$$

It is then easy to calculate that the eigenvalues of $\boldsymbol{\Gamma}_z(0)$ are $\sigma^2$ and $\sigma^2 + L_1^2 + L_2^2$. Since $L_1^2 + L_2^2 > 0$, the largest eigenvalue of $\boldsymbol{\Gamma}_z(0)$ is $\sigma^2 + L_1^2 + L_2^2$. It is also easy to calculate that the eigenvector associated with the largest eigenvalue is exactly proportional to $\boldsymbol{L}$, the factor loading matrix.

### 6.4.2 Approximate Factor Models

The requirement that the covariance matrix of the innovations $\boldsymbol{\epsilon}_t$ being diagonal in Equation (6.16) may be hard to achieve when the dimension $k$ of $\boldsymbol{z}_t$ is large, especially in financial applications in which innovations orthogonal to some known risk factors remain correlated. For this reason, Chamberlain (1983) and Chamberlain and Rothschild (1983) proposed the approximate factor models that allow the innovations $\boldsymbol{\epsilon}_t$ to have a general covariance matrix $\boldsymbol{\Sigma}_\epsilon$. Approximate factor models are commonly used in economics and finance.

Consider an approximate factor model

$$\boldsymbol{z}_t = \boldsymbol{L}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{6.21}$$

where $\boldsymbol{\epsilon}_t$ satisfies (a) $E(\boldsymbol{\epsilon}_t) = \boldsymbol{0}$, (b) $\text{Cov}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}_\epsilon$, and (c) $\text{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_s) = \boldsymbol{0}$ for $t \neq s$. A weakness of this model is that, for a finite $k$, the model is not uniquely identified. For instance, assume that $\boldsymbol{f}_t$ is not serially correlated. Let $\boldsymbol{C}$ be any $k \times k$ orthonormal matrix but $\boldsymbol{I}_k$, and define

$$\boldsymbol{L}^* = \boldsymbol{C}\boldsymbol{L}, \quad \boldsymbol{\epsilon}_t^* = (\boldsymbol{I}_k - \boldsymbol{C})\boldsymbol{L}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t.$$

It is easy to see that we can rewrite the model as

$$\boldsymbol{z}_t = \boldsymbol{L}^*\boldsymbol{f}_t + \boldsymbol{\epsilon}_t^*,$$

which is also an approximate factor model. Forni, et al. (2000) show that the approximate factor model can be identified if the $m$ largest eigenvalues of the covariance matrix of $\boldsymbol{z}_t$ diverge to infinity when $k \to \infty$ and the eigenvalues of the noise covariance matrix remain bounded. Therefore, the approximate factor models in Equation (6.21) is only asymptotically identified.

Another line of research emerges recently in the statistical literature for the approximate factor models in Equation (6.21). See Pan and Yao (2008), Lam et al. (2011), and Lam and Yao (2012). The model considered assumes the same form as Equation (6.21), but requires that (a) $\boldsymbol{L}$ is a $k \times m$ orthonormal matrix so that $\boldsymbol{L}'\boldsymbol{L} = \boldsymbol{I}_m$, (b) $\boldsymbol{\epsilon}_t$ is a white noise series with $E(\boldsymbol{\epsilon}_t) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}_\epsilon$ being a general covariance matrix, (c) $\text{Cov}(\boldsymbol{\epsilon}_t, \boldsymbol{f}_s) = \boldsymbol{0}$ for $s \leq t$, and (d) no linear combinations of $\boldsymbol{f}_t$ are white noise. Some additional assumptions are needed for consistent estimation of $\boldsymbol{L}$ and $\boldsymbol{f}_t$; see Lam, Yao and Bathia (2011) for details. Lam and Yao (2012) investigates the estimation of the number of common factors $m$ when $k$ is large. This approach effectively treats any non-white-noise linear combination as a common factor.

By assumptions, there exists a $k \times (k - m)$ orthonormal matrix $\boldsymbol{U}$ such that $[\boldsymbol{L}, \boldsymbol{U}]$ is a $k \times k$ orthonormal matrix and $\boldsymbol{U}'\boldsymbol{L} = \boldsymbol{0}$. Pre-multiplying Equation (6.21) by $\boldsymbol{U}'$, we have

$$\boldsymbol{U}'\boldsymbol{z}_t = \boldsymbol{U}'\boldsymbol{\epsilon}_t, \tag{6.22}$$

which is a $(k-m)$-dimensional white noise. Based on this white noise property, $\boldsymbol{U}'\boldsymbol{z}_t$ is uncorrelated with $\{\boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-2}, \ldots\}$. In this sense, $\boldsymbol{U}'\boldsymbol{z}_t$ consists of $k - m$ scalar component models of order (0,0), i.e. SCM(0,0) of Chapter **??**. This argument also applies to other factor models. In Tiao and Tsay (1989), search for SCM(0,0) is carried out by canonical correlation analysis between $\boldsymbol{z}_t$ and $\boldsymbol{Z}_{h,t-1} \equiv (\boldsymbol{z}'_{t-1}, \ldots, \boldsymbol{z}'_{t-h})'$ for some selected positive integer $h$. Thus, it considers the rank of $\text{Cov}(\boldsymbol{z}_t, \boldsymbol{Z}_{h,t-1})$ $= [\boldsymbol{\Gamma}_z(1), \ldots, \boldsymbol{\Gamma}_z(h)]$. In Lam and Yao (2012), search for white noise linear combinations is carried out by the eigenanalysis of the matrix $\boldsymbol{G} = \sum_{i=1}^h \boldsymbol{\Gamma}_z(i)\boldsymbol{\Gamma}_z(i)'$, which by construction is non-negative

definite. Thus, the two methods are highly related. On the other hand, Lam and Yao (2012) also considers the case of $k \to \infty$ whereas Tiao and Tsay (1989) assumes $k$ is fixed.
The requirement that no linear combination of $\boldsymbol{f}_t$ is a white noise implies that the orthogonal factor model is not a special case of the approximate factor model of Pan and Yao (2008). This may limit the applicability of the latter model in economics and finance.

### 6.4.3 Diffusion Index Models

In a sequence of papers, Stock and Watson (2002a, b) consider a diffusion index model for prediction. The model is highly related to factor models and can be written as

$$\boldsymbol{z}_t = \boldsymbol{L}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{6.23}$$

$$y_{t+h} = \boldsymbol{\beta}'\boldsymbol{f}_t + \boldsymbol{e}_{t+h}, \tag{6.24}$$

where $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{kt})'$ is a $k$-dimensional stationary time series, which is observed and $E(\boldsymbol{z}_t) = \boldsymbol{0}$, $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{mt})'$ is an $m$-dimensional vector of common factors with $E(\boldsymbol{f}_t) = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{f}_t) = \boldsymbol{I}_m$, $\boldsymbol{L}$ is a $k \times m$ loading matrix, $\boldsymbol{\epsilon}_t$ is a sequence of iid random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}_\epsilon$. Thus, Equation (6.23) is an approximate factor model. The $y_t$ is a scale time series of interest, the positive integer $h$ denotes the forecast horizon, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$ denotes the coefficient vector, and $e_t$ is a sequence of uncorrelated random variables with mean zero and variance $\sigma_e^2$. Equation (6.24) represents the linear equation for $h$-step ahead prediction of $y_{t+h}$ based on the common factor $\boldsymbol{f}_t$. Components of $\boldsymbol{f}_t$ are the diffusion indices. In applications, Equation (6.24) can be extended to include some pre-selected predictors. For instance,

$$y_{t+h} = \boldsymbol{\beta}'\boldsymbol{f}_t + \boldsymbol{\gamma}'\boldsymbol{w}_t + e_{t+h}, \tag{6.25}$$

where $\boldsymbol{w}_t$ consists of some pre-determined predictors such as $y_t$ and $y_{t-1}$. Obviously, a constant term can be added to Equation (6.24) if $E(y_t) \neq 0$.
Since $\boldsymbol{f}_t$ is unobserved, inference concerning the diffusion index model in Equations (6.23) and (6.24) is carried out in two steps. In the first step, one extracts the common factor $\boldsymbol{f}_t$ from the observed high-dimensional series $\boldsymbol{z}_t$. Stock and Watson (2002a, b) apply the principal component analysis to extract $\boldsymbol{f}_t$. Denote the sample principal component by $\hat{\boldsymbol{f}}_t$. The second step is to estimate $\boldsymbol{\beta}$ in Equation (6.24) by the ordinary least square method with $\boldsymbol{f}_t$ replaced by $\hat{\boldsymbol{f}}_t$.
An important contribution of Stock and Watson (2002a) is that the authors show under rather general conditions that (a) $\hat{\boldsymbol{f}}_t$ converges in probability to $\boldsymbol{f}_t$ (up to possible sign changes) and (b) the forecasts derived from Equation (6.24) with $\hat{\boldsymbol{f}}_t$ converge in probability to the mean-square efficient estimates that could be obtained if the common factor $\boldsymbol{f}_t$ was observable as $k \to \infty$ and $T \to \infty$ jointly with a joint growth rates of $k$ and $T$, where $T$ denotes the sample size. In short, these authors successfully extend the traditional PC analysis with finite $k$ to the case in which both $k$ and $T$ approach infinity at some joint growth rates.
The selection of the number of common factors $m$ from $\boldsymbol{z}_t$ has been investigated by several authors. Bai and Ng (2002) extended information criteria such as AIC to account for the effects of increasing $k$ and $T$. Onatski (2009) proposed a test statistic using random matrix theory. Bai (2003) studied limiting properties of the approximate factor models in Equation (6.23). A criticism of the analysis of diffusion index models discussed so far is that the selection of the common factors $\hat{\boldsymbol{f}}_t$ in Equation (6.23) does not make use of the objective function of predicting $y_t$. To overcome this weakness,

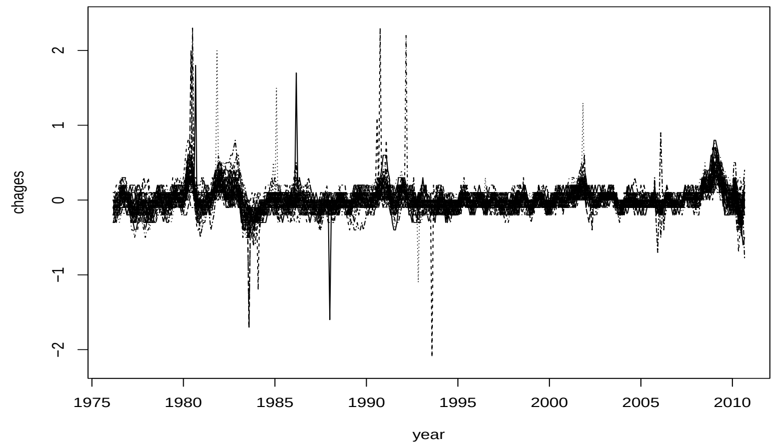


Figure 6.13: Time plots of the first differenced series of the monthly unemployment rates of the U.S. 50 states. The time period is from February 1976 to August 2010. The series were adjusted for outliers.

several alternative methods have been proposed in the recent literature, including using partial least squares and some of its variants. On the other hand, Stock and Watson (2002b) demonstrated that the diffusion index performs well in practice. Finally, Heaton and Solo (2003, 2006) provide some useful insights into the diffusion index models in Equations (6.23) and (6.24).

**Example 6.5**. To illustrate the prediction of diffusion index models, we consider the monthly unemployment rates of the U.S. 50 states from January 1976 to August 2010 for 416 observations. Preliminary analysis indicates the existence of some large outliers and we made some simple adjustments accordingly. The changes are

1. Arizona: Due to a level shift, we subtract 3.3913 from each of the first 92 observations.

2. Louisiana: To adjust for the effect of Hurricane Katrina, we subtracted (6,6,6,1) from the unemployment rates for $t$ from 357 to 360, respectively.

3. Mississippi: To partially remove the effect of Hurricane Katrina, we subtracted (3,2,5,3,2,0.5) from the observed rates for $t$ from 357 to 360, respectively.

Let $\boldsymbol{z}_t$ be the first differenced monthly unemployment rates of the 50 states. Figure 6.13 shows the time plots of the 50 change series in $\boldsymbol{z}_t$. The series are highly related and still contain some aberrant observations.
In our demonstration, we predict $y_{t+1}$ for the first 5 components of $\boldsymbol{z}_t$ using $\boldsymbol{x}_t = (\boldsymbol{z}_t', \boldsymbol{z}_{t-1}', \boldsymbol{z}_{t-2}', \boldsymbol{z}_{t-3}')'$. Thus, we use 200 regressors in the exercise. The five states used as dependent variables are Alabama (AL), Alaska (AK), Arizona (AZ), Arkansas (AR), and California (CA). The forecast origin is 350. In other words, we use the first 350 observations of $\boldsymbol{z}_t$ to construct $\boldsymbol{x}_t$ and perform the principal component analysis on $\boldsymbol{x}_t$ to derive the diffusion indices. The components of $\boldsymbol{x}_t$ is standardized

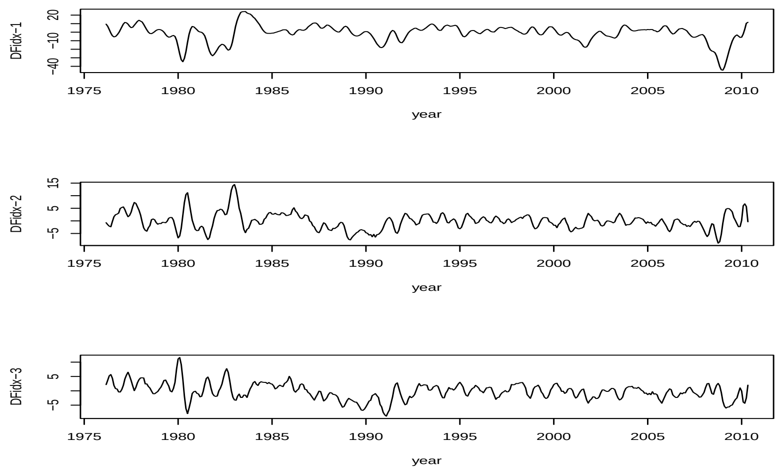


Figure 6.14: Time plots of the first three diffusion indices of the differenced series of monthly state unemployment rates. The indices are based on principal component analysis of the first 350 observations with lags 1 to 4.

individually before the principal component analysis. Denote the diffusion indices by $f_{it}$. We use the prediction equation

$$y_{t+1} = \beta_0 + \sum_{i=1}^{m} \beta_i f_{it} + e_t, \quad t = 5, \ldots, 350 \tag{6.26}$$

to obtain the estimates $\hat{\beta}_i$. The fitted equation is then used to perform one-step ahead predictions for $t = 351, \ldots, 414$. Note that the diffusion indices $f_{it}$ in the forecasting period were obtained by the same loadings used in the estimation period. Figure 6.14 shows the time plots of the first 3 diffusion indices. Based on the PCA, the first diffusion index $f_{1t}$ explains about 50% of the variability in the standardized data. Table 6.9 summarizes the out-of-sample performance of the diffusion index approach for various choices of $m$. From the table, we made the following observations. First, using all 200 regressors did poorly in prediction. As a matter of fact, it is the worst case for all five states considered. This result confirms that over-fitting fares poorly in out-of-sample prediction. Second, no single choice of $m$ outperforms the others. A choice of $m$ between 20 to 50 seems to work well. In applications, one can use out-of-sample prediction like the one shown in Table 6.9 to select $m$. The choice depends on $y_t$, the variable of interest.

**Remark**: The out-of-sample prediction using the diffusion index approach of Stock and Watson (2002a) is obtained by the command `SWfore` of the `MTS` package. Details are in the attached R demonstration.

***R Demonstration***: Forecasting via diffusion index

```
> da=read.table("m-unempstatesAdj.txt",header=T)
```

Table 6.9: Mean Squares of Forecast Errors for 1-step Ahead Out-of-Sample Prediction Using Diffusion Index, where $m$ Denotes the Number of Indices Used. The Forecast Origin is March, 2004. The Numbers Shown are MSE$\times 10^2$.

| | Number of Diffusion Indices $m$ | | | | | |
|---|---|---|---|---|---|---|
| State | 10 | 20 | 30 | 50 | 100 | 200 |
| AL | 1.511 | 1.275 | 1.136 | 1.045 | 1.144 | 2.751 |
| AK | 0.619 | 0.480 | 0.530 | 0.641 | 0.590 | 2.120 |
| AZ | 1.223 | 1.274 | 1.195 | 1.109 | 1.738 | 3.122 |
| AR | 0.559 | 0.466 | 0.495 | 0.473 | 0.546 | 1.278 |
| CA | 0.758 | 0.694 | 0.626 | 0.636 | 0.663 | 0.849 |

```
> dim(da)
[1] 416  50
> source("MTS.R")
> drate=diffM(da)  # First difference
> source("SWfore.R")
> dim(drate)
[1] 415  50
> y=drate[5:415,1]  # First y_t series (Alabama)
> length(y)
[1] 411
> x=cbind(drate[4:414,],drate[3:413,],drate[2:412,],drate[1:411,]) #z_t series
> dim(x)
[1] 411 200
> m1=SWfore(y,x,350,10)
MSE of out-of-sample forecasts:  0.01510996
> m1=SWfore(y,x,350,20)
MSE of out-of-sample forecasts:  0.01274754
> m1=SWfore(y,x,350,30)
MSE of out-of-sample forecasts:  0.01136177
> m1=SWfore(y,x,350,50)
MSE of out-of-sample forecasts:  0.01044645
> m1=SWfore(drate[5:415,1],x,350,200)
MSE of out-of-sample forecasts:  0.02750807.
```

### 6.4.4 Dynamic Factor Models

In our discussion of factor models so far, we do not consider explicitly the dynamic dependence of the common factors $\boldsymbol{f}_t$. Forni et al. (2000, 2004, 2005) proposed the dynamic factor model

$$\boldsymbol{z}_t = \boldsymbol{L}(B)\boldsymbol{u}_t + \boldsymbol{\epsilon}_t, \tag{6.27}$$

where $\boldsymbol{\epsilon}_t$ is a white noise series with mean zero and covariance matrix $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{u}_t$ is an $m$-dimensional process of orthonormal white noise, and $\boldsymbol{L}(B) = \boldsymbol{L}_0 + \boldsymbol{L}_1 B + \cdots + \boldsymbol{L}_r B^r$ is a $k \times m$ matrix polynomial of order $r$ (might be infinity), and $B$ is the backshift operator such that $B\boldsymbol{u}_t = \boldsymbol{u}_{t-1}$. By orthonormal white noise we mean that $\boldsymbol{u}_t$ satisfies $E(\boldsymbol{u}_t) = \boldsymbol{0}$, $\mathrm{Cov}(\boldsymbol{u}_t) = \boldsymbol{I}_m$ and $\boldsymbol{u}_t$ is serially uncorrelated. Let $\boldsymbol{L}(B) = [L_{ij}(B)]$. Coefficients of each polynomial $L_{ij}(B)$ are assumed to be square summable. Let $\boldsymbol{c}_t = \boldsymbol{L}(B)\boldsymbol{u}_t = (c_{1t}, \ldots, c_{kt})'$, the model becomes

$$\boldsymbol{z}_t = \boldsymbol{c}_t + \boldsymbol{\epsilon}_t. \tag{6.28}$$

Forni et al. (2000) refers to $c_{it}$ and $\epsilon_{it}$ as the common component and idiosyncratic component of $z_{it}$, respectively. The $\boldsymbol{u}_t$ process is referred to as the common shocks.

Assuming that $\boldsymbol{z}_t$ of Equation (6.27) is stationary, Forni et al. (2000) employed the spectral density matrix of $\boldsymbol{z}_t$ to propose an estimation method for the common components $\boldsymbol{c}_t$ when both $k$ and the sample size $T$ go to infinity in some given path. Under some regularity conditions, the authors show that the estimates converge to the true $\boldsymbol{c}_t$ in mean square. Details are given in Forni et al. (2000).

If we assume that the common factor $\boldsymbol{f}_t$ of the approximate factor model in Equation (6.21) follows a stationary VARMA$(p, q)$ model

$$\boldsymbol{\phi}(B)\boldsymbol{f}_t = \boldsymbol{\theta}(B)\boldsymbol{u}_t,$$

where $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are $m \times m$ AR and MA matrix polynomial, respectively. Then, the model becomes

$$\boldsymbol{z}_t = \boldsymbol{L}[\boldsymbol{\phi}(B)]^{-1}\boldsymbol{\theta}(B)\boldsymbol{u}_t + \boldsymbol{\epsilon}_t.$$

Therefore, we have $\boldsymbol{L}(B) = \boldsymbol{L}[\boldsymbol{\phi}(B)]^{-1}\boldsymbol{\theta}(B)$. This connection between the two models provides an alternative approach to estimate the dynamic factor models. Specifically, if the common shock $\boldsymbol{u}_t$ is of interest, it can be estimated by first building an approximate factor model for $\boldsymbol{z}_t$, then modeling the estimated latent process $\hat{\boldsymbol{f}}_t$ via VARMA models. The latter can be done by the methods discussed in the previous chapters.

### 6.4.5 Constrained Factor Models

Empirical applications of the factor models in Equations (6.16) and (6.21) often show that the estimated loading matrix exhibits certain characteristic patterns. For instance, consider the monthly log returns of ten U.S. stocks in Table 6.7. The fitted loading matrices via either PCA or MLE show that for each column of the loading matrices the loading weights are similar for companies in the same industrial sector. See the estimated loadings in Table 6.8. Motivated by the observation and in preference of parsimonious loading matrices for ease in interpretation, Tsai and Tsay (2010) propose a constrained factor model that can explicitly describe the observed patterns. For simplicity, assume that $E(\boldsymbol{z}_t) = \boldsymbol{0}$. A constrained factor model can be written as

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{\omega}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{6.29}$$

where $\boldsymbol{f}_t$ and $\boldsymbol{\epsilon}_t$ are as defined before, $\boldsymbol{H}$ is a known $k \times r$ matrix and $\boldsymbol{\omega}$ is a $r \times m$ matrix of unknown parameters. The matrix $\boldsymbol{H}$ is a constraint matrix with each column giving rise to a specific constraint. For instance, consider the ten stocks of Example 6.4. These ten stocks belong to three industrial sectors so that we can define $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3]$ with $\boldsymbol{h}_i$ denoting the $i$th sector. To demonstrate, let

$$\begin{aligned} \boldsymbol{h}_1 &= (1,1,1,1,0,0,0,0,0,0)' \\ \boldsymbol{h}_2 &= (0,0,0,0,1,1,1,0,0,0)' \\ \boldsymbol{h}_3 &= (0,0,0,0,0,0,0,1,1,1)'. \end{aligned} \tag{6.30}$$

Then, $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, and $\boldsymbol{h}_3$ represent the sector of semiconductor, pharmaceutical, and investment bank, respectively. In practice, $\boldsymbol{H}$ is flexible and can be specified using prior information or economic theory.

The number of parameters in the loading matrix of the constrained factor model in Equation (6.29) is $r \times m$ whereas that of the unconstrained factor model is $k \times m$. When $r$ is much smaller than $k$, the constrained factor model can be much more parsimonious. Under the same assumptions as the traditional factor models, we have

$$\begin{aligned} \boldsymbol{\Gamma}_z(0) &= \boldsymbol{H}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{H}' + \boldsymbol{\Sigma}_\epsilon, \\ \boldsymbol{\Gamma}_z(\ell) &= \boldsymbol{H}\boldsymbol{\omega}\boldsymbol{\Gamma}_f(\ell)\boldsymbol{\omega}'\boldsymbol{H}', \end{aligned}$$

for the constrained factor model. These properties can be used to estimate the constrained factor models. In general, the constrained model can also be estimated by either the least squares method or maximum likelihood method. The maximum likelihood method assumes normality and requires the additional constraint $\widehat{\boldsymbol{\omega}}'\boldsymbol{H}'\widehat{\boldsymbol{\Sigma}}_\epsilon\boldsymbol{H}\widehat{\boldsymbol{\omega}}$ is a diagonal matrix. There is no closed-form solution and an iterated procedure is used. Details are given in Tsai and Tsay (2010). Here we briefly discuss the least squares method.

Write the data of constrained factor model in Equation (6.29) as

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{\omega}'\boldsymbol{H}' + \boldsymbol{\epsilon}, \tag{6.31}$$

where $\boldsymbol{Z} = [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T]$ is a $T \times k$ data matrix, $\boldsymbol{F} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T]'$ is the $T \times m$ matrix of common factors, $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T]'$ and $T$ is the sample size. Let $\mathrm{tr}(\boldsymbol{A})$ be the trace of matrix $\boldsymbol{A}$. The least squares approach to estimate $\boldsymbol{F}$ and $\boldsymbol{\omega}$ is to minimize the objective function

$$\ell(\boldsymbol{F}, \boldsymbol{\omega}) = \mathrm{tr}[(\boldsymbol{Z} - \boldsymbol{F}\boldsymbol{\omega}'\boldsymbol{H}')(\boldsymbol{Z} - \boldsymbol{F}\boldsymbol{\omega}'\boldsymbol{H}')'], \tag{6.32}$$

subject to the constraint $\boldsymbol{F}\boldsymbol{F} = T\boldsymbol{I}_m$. Using least squares theory, we have $\widehat{\boldsymbol{\omega}} = \boldsymbol{T}^{-1}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\boldsymbol{F}$. Plugging $\widehat{\boldsymbol{\omega}}$ into Equation (6.32) and using $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$, we obtain the concentrated function

$$\ell(\boldsymbol{F}) = \mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}') - \boldsymbol{T}^{-1}\mathrm{tr}[\boldsymbol{F}'\boldsymbol{Z}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\boldsymbol{F}].$$

This objective function is minimized when the second term is maximized. Applying theorem 6 of Magnus and Neudecker (1999, p. 205) or proposition A.4 of Lütkepohl (2005, p. 672), we have $\widehat{\boldsymbol{F}} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m]$, where $\boldsymbol{g}_i$ is an eigenvector of the $i$th largest eigenvalue $\lambda_i$ of $\boldsymbol{Z}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'$. In practice, the eigenvectors are normalized so that $\widehat{\boldsymbol{F}'}\widehat{\boldsymbol{F}} = T\boldsymbol{I}_m$. The corresponding estimate of $\boldsymbol{\omega}$ becomes $\widehat{\boldsymbol{\omega}} = T^{-1}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\widehat{\boldsymbol{F}}$. Finally, the covariance matrix of the noises is estimated by $\widehat{\boldsymbol{\Sigma}}_\epsilon = T^{-1}\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{H}\widehat{\boldsymbol{\omega}}\widehat{\boldsymbol{\omega}}'\boldsymbol{H}'$. Properties of the least squares estimates are studied in Tsai and Tsay (2010) and the references therein. Tsai and Tsay (2010) also consider test statistics for checking constraints in a factor model and a partially constrained factor model.

**Example 6.6**. Consider again the monthly log returns of ten U.S. companies given in Table 6.7. We apply the constrained factor model to the returns using the constraint matrix $\boldsymbol{H}$ of Equation (6.30) using the least squares method. The result is summarized in Table 6.10. For ease in comparison, Table 6.10 also reports the result of the principal component approach to the orthogonal factor model as shown in Table 6.8. From the table, we make the following observations. First, in this particular application, we have $r = m = 3$ and $k = 10$. The constrained factor model uses 9

Table 6.10: Estimation Results of Constrained and Orthogonal Factor Models for the Monthly Log Returns of Ten U.S. Companies given in Table 6.7, where $\boldsymbol{L}_i$ is the $i$th column of $\boldsymbol{L}$ and $\Sigma_{\epsilon,i}$ is the $(i,i)$th element of $\boldsymbol{\Sigma}_\epsilon$.

| Stock | Constrained Model: $\widehat{\boldsymbol{L}} = \boldsymbol{H}\widehat{\boldsymbol{\omega}}$ | | | | Orthogonal Model: PCA | | | |
|---|---|---|---|---|---|---|---|---|
| Tick | $\boldsymbol{L}_1$ | $\boldsymbol{L}_2$ | $\boldsymbol{L}_3$ | $\Sigma_{\epsilon,i}$ | $\boldsymbol{L}_1$ | $\boldsymbol{L}_2$ | $\boldsymbol{L}_3$ | $\Sigma_{\epsilon,i}$ |
| TXN | 0.761 | 0.256 | 0.269 | 0.283 | 0.788 | 0.198 | 0.320 | 0.237 |
| MU | 0.761 | 0.256 | 0.269 | 0.283 | 0.671 | 0.361 | 0.289 | 0.336 |
| INTC | 0.761 | 0.256 | 0.269 | 0.283 | 0.789 | 0.183 | 0.333 | 0.232 |
| TSM | 0.761 | 0.256 | 0.269 | 0.283 | 0.802 | 0.270 | 0.159 | 0.258 |
| PFE | 0.444 | -0.675 | 0.101 | 0.337 | 0.491 | -0.643 | -0.027 | 0.345 |
| MRK | 0.444 | -0.675 | 0.101 | 0.337 | 0.402 | -0.689 | 0.226 | 0.312 |
| LLY | 0.444 | -0.675 | 0.101 | 0.337 | 0.448 | -0.698 | 0.061 | 0.309 |
| JPM | 0.738 | 0.055 | -0.431 | 0.267 | 0.724 | 0.020 | -0.345 | 0.357 |
| MS | 0.738 | 0.055 | -0.431 | 0.267 | 0.755 | 0.053 | -0.425 | 0.246 |
| GS | 0.738 | 0.055 | -0.431 | 0.267 | 0.745 | 0.122 | -0.498 | 0.182 |
| e.v. | 4.576 | 1.650 | 0.883 | | 4.626 | 1.683 | 0.932 | |
| | Variability explained: 70.6% | | | | Variability explained: 72.4% | | | |

parameters in the matrix $\boldsymbol{\omega}$. Its loading matrix is obtained by $\widehat{\boldsymbol{L}} = \boldsymbol{H}\widehat{\boldsymbol{\omega}}$. Therefore, the loading matrix applies the same weights to each stock in a given sector. This seems reasonable as the returns are standardized to have unit variance. On the other hand, the orthogonal factor model has 30 parameters in the loading matrix. The weights are close, but not identical for each stock in the same sector. Second, with the constraints, the estimated 3-factor model explains about 70.6% of the total variability. This is very close to the unconstrained 3-factor model, which explains 72.4% of the total variability. This confirms that the constraints are acceptable. Third, with the constraints, the first common factor $f_{1t}$ represents the market factor, which is a weighted average of the three sectors employed. The second factor $f_{2t}$ is essentially a weighted difference between the semiconductor and pharmaceutical sector. The third common factor $f_{3t}$ represents the difference between investment banks and the other two sectors.

**Discussion**: At the first glance, the constraint matrix $\boldsymbol{H}$ appears to be awkward. However, in applications, $\boldsymbol{H}$ is flexible and through which one can incorporate into the analysis substantive prior knowledge of the problem at hand. For instance, it is well-known that the term structure of interest rates can be approximately described by trend, slope, and curvature. This knowledge can be used to specify $\boldsymbol{H}$. For simplicity, suppose $\boldsymbol{z}_t = (z_{1t}, z_{2t}, z_{3t})'$ with components representing interest rates with short, intermediate, and long maturities, respectively. Then, one can use $\boldsymbol{h}_1 = (1,1,1)'$, $\boldsymbol{h}_2 = (-1,0,1)'$ and $\boldsymbol{h}_3 = (1,-2,1)'$ to represent trend, slope, and curvature, respectively. The choices of $\boldsymbol{H}$ become more flexible when $k$ is large.

**Remark**. The least squares estimation of constrained factor models can be carried out via the command `hfactor` in the `MTS` package.

***R Demonstration***: Constrained factor models. Output edited.

```
> da=read.table("m-tenstocks.txt",header=T)
> rtn=log(da[,2:11]+1)  # compute log returns
> h1=c(1,1,1,1,rep(0,6))  # specify the constraints
> h2=c(0,0,0,0,1,1,1,0,0,0)
> h3=c(rep(0,7),1,1,1)
> H=cbind(h1,h2,h3)
> m1=hfactor(rtn,H,3)
[1] "Data are individually standardized"
[1] "First m eigenvalues of the correlation matrix:"
[1] 4.6256602 1.6827255 0.9320882
[1] "Variability explained: "
[1] 0.7240474
[1] "Loadings:"
        [,1]    [,2]    [,3]
 [1,] -0.368 -0.1532 -0.3331
 [2,] -0.313 -0.2792 -0.3000
 [3,] -0.368 -0.1419 -0.3465
 [4,] -0.374 -0.2090 -0.1649
 [5,] -0.229  0.4978  0.0278
 [6,] -0.188  0.5333 -0.2354
 [7,] -0.209  0.5401 -0.0632
 [8,] -0.338 -0.0153  0.3586
 [9,] -0.352 -0.0411  0.4421
[10,] -0.348 -0.0946  0.5173
[1] "eigenvalues of constrained part:"
[1] 4.576 1.650 0.883
[1] "Omega-Hat"
      [,1]    [,2]   [,3]
[1,] 0.761  0.2556  0.269
[2,] 0.444 -0.6752  0.101
[3,] 0.738  0.0547 -0.431
[1] "Variation explained by the constrained factors:"
[1] 0.7055665
[1] "H*Omega: constrained loadings"
       [,1]    [,2]   [,3]
 [1,] 0.761  0.2556  0.269
 [2,] 0.761  0.2556  0.269
 [3,] 0.761  0.2556  0.269
 [4,] 0.761  0.2556  0.269
 [5,] 0.444 -0.6752  0.101
 [6,] 0.444 -0.6752  0.101
 [7,] 0.444 -0.6752  0.101
 [8,] 0.738  0.0547 -0.431
 [9,] 0.738  0.0547 -0.431
[10,] 0.738  0.0547 -0.431
[1] "Diagonal elements of Sigma_epsilon:"
  TXN    MU  INTC   TSM   PFE   MRK   LLY   JPM    MS    GS
0.283 0.283 0.283 0.283 0.337 0.337 0.337 0.267 0.267 0.267
[1] "eigenvalues of Sigma_epsilon:"
```

```
[1]  0.7632  0.6350  0.4539  0.3968  0.3741  0.2474  0.2051
[8]  0.0262 -0.0670 -0.0902
```