STAT 626: Review: Multivariate Time Series Analysis (§5.8)

1. ~~**Problem 3.42: Normal equations, Invertibility, etc.**~~

2. **How to Model Several TS Simultaneously?**

3. **VMA(1) Models:** $X_t = W_t + \Theta W_{t-1} = (I + \Theta B)W_t, \quad W_t \sim \mathbf{WN}(0, \Sigma)$
   ACF, PACF, Invertibility,....?

4. **VAR(1) Models:** $X_t = \Phi X_{t-1} + W_t, , (I - \Phi B)X_t = W_t, \quad W_t \sim \mathbf{WN}(0, \Sigma)$
   ACF, PACF, Causality, ...?

5. **Writing AR(p) as p-dim. VAR(1)**

6. **VARMA(1,1)?**

7. **Interrelatioships Among Several TS: Granger Causality**

8. **Jointly Stationary TS: Focus on VAR(1)**

9. **Co-Ingetrated TS: VAR(1) with Unit Roots**

10. **Multivariate ARMAX Models (§5.8)**

STAT 626: Outline Lecture 19
ARCH-GARCH Models (§8.1)

1. **WN $\Rightarrow$ ARMA, ARIMA, SARIMA, ARCH/GARCH, ...**

2. **Taking Care of Time-Varying Variances:** σ_t^2

3. **Time Series Decomposition:** $x_t = \mu_t + \sigma_t \varepsilon_t, \quad$ **Var**$(\sigma_t \varepsilon_t) = \sigma_t^2$.

4. **How to Model Time-Varying Variances?**
   Recall that Squared Residuals r_t^2 are Reasonable "Estimates" of σ_t^2:

$$r_t^2 \approx \sigma_t^2.$$

5. **Often r_t^2's appear more correlated than r_t's (Granger, 1970's).**

6. **AutoRegressive Conditionally Heteroscedastic (ARCH) Models:**(Engle, 1982)

$$r_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2.$$

   AR Models for Squared Residuals r_t^2.

   This point of view is helpful in using the ACF and PACF of the series r_t^2 to identify the orders of the ARCH(p) models.

7. **Generalized ARCH (GARCH) Models**

   ARMA Models for Squared Residuals r_t^2.

Unit-Root Test and Random Walks

8. **Random Walk vs AR(1):** $x_t = \phi x_{t-1} + w_t$,

$$H_0 : \phi = 1 \quad \text{vs} \quad H_1 : |\phi| < 1.$$

9. **Unit-Root Tests: DF, ADF, PP**.

10. **Why Unit-Root Test is Important in Economics and Finance?**

L. Bachelier Dissertation (1900).

Random Walk Hypothesis,

Efficient Market Hypothesis:

The weak form: All information about market prices is already reflected in the current stock price.

The strong form: All publicly available information about a company is already reflected in its stock price.

I. A Random Walk Down Wall Street, by Burton G. Malkiel

II. A Non-Random Walk Down Wall Street, by Andrew W. Lo & A. Craig MacKinlay

Books By Peter Bernstein:

III. Capital Ideas: The Improbable Origins of Modern Wall Street, (Free Press), 1991.

IV. Against the Gods: The Remarkable Story of Risk, (John Wiley & Son), 1996,
Story of (Random Walk) Brownian Motion and how it Entered the World of Finance.

STAT 626: Time Series Analysis; Some Guidelines for Presenting Group Projects in Zoom and Beyond

In the presentations and writing reports, please keep the following in mind:

5-min. Presentation and 1-page Written Report

1. Introduce Your Group: List members of your group, their majors, roles and responsibilities (computation, programming, writing, theory,...). Divide the task of presentation (in the 2nd and Final) among the group so that everybody gets the chance to present.
2. The Goal of the Project: State the primary goal using plain English.
3. Introduce Your Data: Explain the context and background story of your data so that your audience develops a good feel for the data and the subject they are coming from.
4. Plot the Data and summarize your preliminary findings: trend, cycle, volatility, etc. **This first report should serve as the cover page of the next two reports with minor changes.**
5. Your group can always make an appointment and discuss with me if you got questions.

For Second Presentation and 5-page Report

6. Transform the Data to Stationarity using regression (follow Example 3.5 in the text as much as possible); Explain differencing, log-transform,.... if used.
7. ACF and PACF Plots: Use correlogram and partial correlogram to formulate ARMA(p, q) models for the "stationary" data. If in doubt, choose from AR models, these are simple to estimate, interpret and predict.
8. Fit and Forecast: Estimate the model parameters using simple-minded methods like the least squares, Yule-Walker estimates, etc.
9. Diagnostic: Check the residuals to see if they are white noise.
10. You may want to consult the first three chapters of the text for notation, terminologies and ideas.

For the Final Presentation and 10-page Report

11. Incorporate as much as feasible the comments/questions came up during the first two presentations. Use model building techniques from Chap.5
12. GRADE: Your presentation and written reports are graded based on their qualities, and the level of interest/question/enthusiasm they generates from other students.

GOOD LUCK

16

STAT 626: Outline of Lectures ~~20 and 21~~

Forecasting (§4.4, 5.1)

1. **Best Linear Prediction for Stationary ARMA Models**

2. **Integrated (ARIMA) Models (§5.1)**

3. **Best Linear Prediction for Nonstationary ARIMA Models**

4. **Given the time series data $x_1, \ldots, x_n$: What is a good way to forecast the next future value x_{n+1}?**

5. **Prediction, Error Variance (P_{n+1}^n) and Forecast Interval:**

$$x_{n+1}^n \pm 1.96\sqrt{P_{n+1}^n}$$

Who do you think said the following:?
"It is difficult to make predictions, especially about the future"

Fun Reading

6. Graham Southorn (2016). **Great expectations: The past, present and future of prediction.** *Significance*, April Issue.

7. Philip Tetlock and Dan Gardner (2015). **Superforcasting: The Art and Science of Prediction.**

8. Michael Abramowicz (2008). **Predictocracy: Market Mechanisms for Public and Private Decision Making.**

   Predicting the future is serious business for virtually all public and private institutions, for they must often make important decisions based on such predictions. This visionary book explores how institutions from legislatures to corporations might improve their predictions and arrive at better decisions by means of prediction markets, a promising new tool with virtually unlimited potential applications.

9. James Surowiecki (2005). **The Wisdom of Crowds: Why the Many are Smarter than the Few**

FORECASTING : Time keeps on slipping, slipping into the future...(Steve Miller Band)

How is statistical or scientific forecast different from that of a **fortune teller or psychic?**

Given the time series data $x_1, \ldots, x_n$: What is a good way to forecast the next future value x_{n+1}?

For the moment pretend the x_i are i.i.d. Then, their sample mean is the "best" predictor. Why?

In what sense is it the "best" predictor?

$$\boldsymbol{x}_t = \Gamma \boldsymbol{u}_t + \sum_{j=1}^{p} \Phi_j \boldsymbol{x}_{t-j} + \boldsymbol{w}_t, \tag{5.86}$$

where Γ is a $p \times r$ parameter matrix. The X in ARX refers to the exogenous vector process we have denoted here by $\boldsymbol{u}_t$. The introduction of exogenous variables through replacing $\boldsymbol{\alpha}$ by $\Gamma \boldsymbol{u}_t$ does not present any special problems in making inferences and we will often drop the X for being superfluous.

Example 5.10 Pollution, Weather, and Mortality

For example, for the three-dimensional series composed of cardiovascular mortality x_{t1}, temperature x_{t2}, and particulate levels x_{t3}, introduced in Example 2.2, take $\boldsymbol{x}_t = (x_{t1}, x_{t2}, x_{t3})'$ as a vector of dimension $k = 3$. We might envision dynamic relations among the three series defined as the first order relation,

$$x_{t1} = \alpha_1 + \beta_1 t + \phi_{11} x_{t-1,1} + \phi_{12} x_{t-1,2} + \phi_{13} x_{t-1,3} + w_{t1},$$

which expresses the current value of mortality as a linear combination of trend and its immediate past value and the past values of temperature and particulate levels. Similarly,

$$x_{t2} = \alpha_2 + \beta_2 t + \phi_{21} x_{t-1,1} + \phi_{22} x_{t-1,2} + \phi_{23} x_{t-1,3} + w_{t2}$$

and

$$x_{t3} = \alpha_3 + \beta_3 t + \phi_{31} x_{t-1,1} + \phi_{32} x_{t-1,2} + \phi_{33} x_{t-1,3} + w_{t3}$$

express the dependence of temperature and particulate levels on the other series. Of course, methods for the preliminary identification of these models exist, and we will discuss these methods shortly. The model in the form of (5.86) is

$$\boldsymbol{x}_t = \Gamma \boldsymbol{u}_t + \Phi \boldsymbol{x}_{t-1} + \boldsymbol{w}_t,$$

where, in obvious notation, $\Gamma = [\boldsymbol{\alpha} \mid \boldsymbol{\beta}]$ is 3×2 and $\boldsymbol{u}_t = (1, t)'$ is 2×1.

Throughout much of this section we will use the R package `vars` to fit vector AR models via least squares. For this particular example, we have (partial output shown):

```
library(vars)
x = cbind(cmort, tempr, part)
summary(VAR(x, p=1, type="both"))  # "both" fits constant + trend
```

```
Estimation results for equation cmort:  (other equations not shown)
   cmort = cmort.l1 + tempr.l1 + part.l1 + const + trend

             Estimate Std. Error t value Pr(>|t|)
   cmort.l1  0.464824   0.036729  12.656  < 2e-16 ***
   tempr.l1 -0.360888   0.032188 -11.212  < 2e-16 ***
   part.l1   0.099415   0.019178   5.184 3.16e-07 ***
   const    73.227292   4.834004  15.148  < 2e-16 ***
   trend    -0.014459   0.001978  -7.308 1.07e-12 ***
```

```
Residual standard error: 5.583 on 502 degrees of freedom
Multiple R-Squared: 0.6908,     Adjusted R-squared: 0.6883
F-statistic: 280.3 on 4 and 502 DF,  p-value: < 2.2e-16

Covariance matrix of residuals:
       cmort  tempr   part
cmort 31.172  5.975  16.65
tempr  5.975 40.965  42.32
part  16.654 42.323 144.26
```

Note that t here is `time(cmort)`, which is $1970 + \Delta\,(t-1)$, where $\Delta = 1/52$, for $t = 1, ..., 508$, ending at 1979.75. For this particular case, we obtain

$$\widehat{\boldsymbol{\alpha}} = (73.23, 67.59, 67.46)', \qquad \widehat{\boldsymbol{\beta}} = (-0.014, -0.007, -0.005)',$$

$$\widehat{\Phi} = \begin{pmatrix} .46(.04) & -.36(.03) & .10(.02) \\ -.24(.04) & .49(.04) & -.13(.02) \\ -.12(.08) & -.48(.07) & .58(.04) \end{pmatrix},$$

where the standard errors, computed as in (5.80), are given in parentheses. The estimate of Σ_w seen to be

$$\widehat{\Sigma}_w = \begin{pmatrix} 31.17 & 5.98 & 16.65 \\ 5.98 & 40.965 & 42.32 \\ 16.65 & 42.32 & 144.26 \end{pmatrix}.$$

For the vector $(x_{t1}, x_{t2}, x_{t3}) = (M_t, T_t, P_t)$, with M_t, T_t and P_t denoting mortality, temperature, and particulate level, respectively, we obtain the prediction equation for mortality,

$$\widehat{M_t} = 73.23 - .014t + .46M_{t-1} - .36T_{t-1} + .10P_{t-1}.$$

Comparing observed and predicted mortality with this model leads to an R^2 of about .69.

It is easy to extend the VAR(1) process to higher orders, VAR(p). To do this, we use the notation of (5.77) and write the vector of regressors as

$$\boldsymbol{z}_t = (1, \boldsymbol{x}'_{t-1}, \boldsymbol{x}'_{t-2}, \ldots \boldsymbol{x}'_{t-p})'$$

and the regression matrix as $\mathcal{B} = (\boldsymbol{\alpha}, \Phi_1, \Phi_2, \ldots, \Phi_p)$. Then, this regression model can be written as

$$\boldsymbol{x}_t = \boldsymbol{\alpha} + \sum_{j=1}^{p} \Phi_j \boldsymbol{x}_{t-j} + \boldsymbol{w}_t \tag{5.87}$$

for $t = p+1, \ldots, n$. The $k \times k$ error sum of products matrix becomes

$$SSE = \sum_{t=p+1}^{n} (\boldsymbol{x}_t - \mathcal{B}\boldsymbol{z}_t)(\boldsymbol{x}_t - \mathcal{B}\boldsymbol{z}_t)', \tag{5.88}$$

so that the conditional maximum likelihood estimator for the error covariance matrix Σ_w is

$$\widehat{\Sigma}_w = SSE/(n-p), \tag{5.89}$$

as in the multivariate regression case, except now only $n-p$ residuals exist in (5.88). For the multivariate case, we have found that the Schwarz criterion

$$\text{BIC} = \log|\widehat{\Sigma}_w| + k^2 p \ln n/n, \tag{5.90}$$

2

gives more reasonable classifications than either AIC or corrected version AICc. The result is consistent with those reported in simulations by Lütkepohl (1985).

Example 5.11 Pollution, Weather, and Mortality (cont)

We used the R package first to select a VAR(p) model and then fit the model. The selection criteria used in the package are AIC, Hannan-Quinn (`HQ`; Hannan & Quinn, 1979), BIC (`SC`), and Final Prediction Error (`FPE`). The Hannan-Quinn procedure is similar to BIC, but with $\ln n$ replaced by $2\ln(\ln(n))$ in the penalty term. FPE finds the model that minimizes the approximate mean squared one-step-ahead prediction error (see Akaike, 1969 for details); it is rarely used.

```
VARselect(x, lag.max=10, type="both")
```

```
$selection          BIC
AIC(n)  HQ(n)  SC(n) FPE(n)
     9      5      2      9
$criteria
                   1           2           3           4
AIC(n)      11.73780    11.30185    11.26788    11.23030
HQ(n)       11.78758    11.38149    11.37738    11.36967
SC(n)       11.86463    11.50477    11.54689    11.58541
FPE(n) 125216.91717 80972.28678 78268.19568 75383.73647
                  5           6           7           8
AIC(n)     11.17634    11.15266    11.15247    11.12878
HQ(n)      11.34557    11.35176    11.38144    11.38760
SC(n)      11.60755    11.65996    11.73587    11.78827
FPE(n) 71426.10041 69758.25113 69749.89175 68122.40518
                  9          10
AIC(n)     11.11915    11.12019
HQ(n)      11.40784    11.43874
SC(n)      11.85473    11.93187
FPE(n) 67476.96374 67556.45243
```

Note that BIC picks the order $p = 2$ model while AIC and FPE pick an order $p = 9$ model and Hannan-Quinn selects an order $p = 5$ model.

Fitting the model selected by BIC we obtain

$$\widehat{\boldsymbol{\alpha}} = (56.1, 49.9, 59.6)', \qquad \widehat{\boldsymbol{\beta}} = (-0.011, -0.005, -0.008)',$$

$$\widehat{\Phi}_1 = \begin{pmatrix} .30(.04) & -.20(.04) & .04(.02) \\ -.11(.05) & .26(.05) & -.05(.03) \\ .08(.09) & -.39(.09) & .39(.05) \end{pmatrix},$$

$$\widehat{\Phi}_2 = \begin{pmatrix} .28(.04) & -.08(.03) & .07(.03) \\ -.04(.05) & .36(.05) & -.10(.03) \\ -.33(.09) & .05(.09) & .38(.05) \end{pmatrix},$$

where the standard errors are given in parentheses. The estimate of Σ_w is

$$\widehat{\Sigma}_w = \begin{pmatrix} 28.03 & 7.08 & 16.33 \\ 7.08 & 37.63 & 40.88 \\ 16.33 & 40.88 & 123.45 \end{pmatrix}.$$

To fit the model using the `vars` package use the following line (partial results displayed):

```
summary(fit <- VAR(x, p=2, type="both"))
```

```
  cmort = cmort.l1 + tempr.l1 + part.l1 + cmort.l2 + tempr.l2
           + part.l2 + const + trend

             Estimate Std. Error t value Pr(>|t|)
  cmort.l1   0.297059   0.043734   6.792 3.15e-11 ***
  tempr.l1 -0.199510   0.044274  -4.506 8.23e-06 ***
  part.l1    0.042523   0.024034   1.769  0.07745 .
  cmort.l2   0.276194   0.041938   6.586 1.15e-10 ***
  tempr.l2 -0.079337   0.044679  -1.776  0.07639 .
  part.l2    0.068082   0.025286   2.692  0.00733 **
  const     56.098652   5.916618   9.482  < 2e-16 ***
  trend     -0.011042   0.001992  -5.543 4.84e-08 ***

  Covariance matrix of residuals:
         cmort  tempr   part
  cmort 28.034  7.076  16.33
  tempr  7.076 37.627  40.88
  part  16.325 40.880 123.45
```

Using the notation of the previous example, the prediction model for cardiovascular mortality is estimated to be

$$\widehat{M}_t = 56 - .01t + .3M_{t-1} - .2T_{t-1} + .04P_{t-1} + .28M_{t-2} - .08T_{t-2} + .07P_{t-2}.$$

To examine the residuals, we can plot the cross-correlations of the residuals and examine the multivariate version of the Q-test as follows:

```
acf(resid(fit), 52)
serial.test(fit, lags.pt=12, type="PT.adjusted")
```

```
    Portmanteau Test (adjusted)
    data:  Residuals of VAR object fit
    Chi-squared = 162.3502, df = 90, p-value = 4.602e-06
```

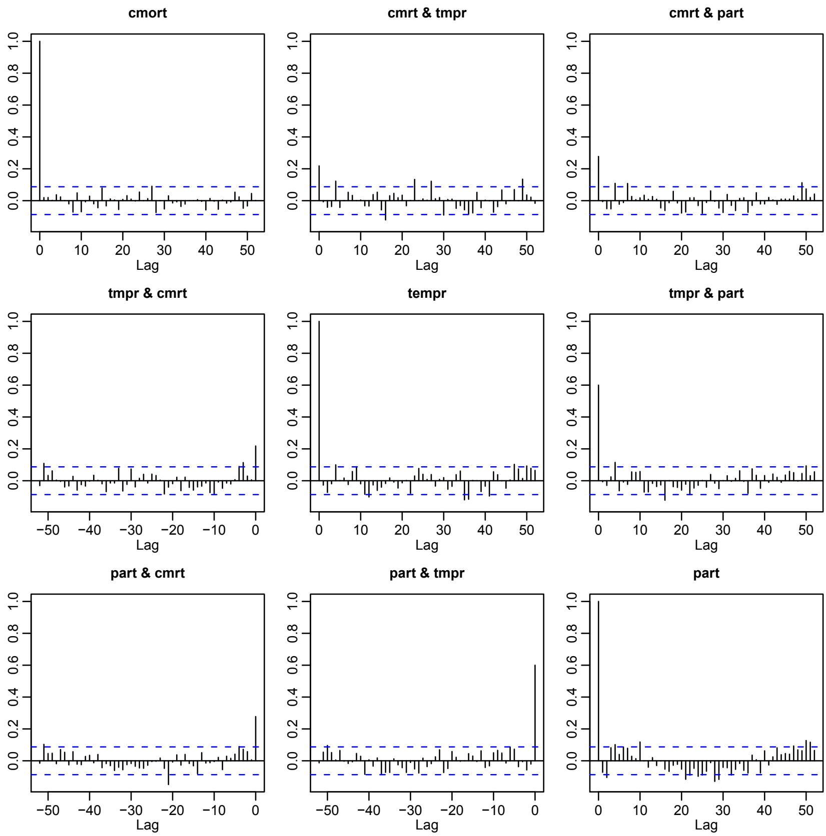

Fig. 5.13. ACFs (diagonals) and CCFs (off-diagonals) for the residuals of the three-dimensional VAR(2) fit to the LA mortality – pollution data set. On the off-diagonals, the second-named series is the one that leads.

The cross-correlation matrix is shown in Figure 5.13. The figure shows the ACFs of the individual residual series along the diagonal. For example, the first diagonal graph is the ACF of $M_t - \widehat{M_t}$, and so on. The off diagonals display the CCFs between pairs of residual series. If the title of the off-diagonal plot is `x & y`, then `y` leads in the graphic; that is, on the upper-diagonal, the plot shows `corr[x(t+Lag), y(t)]` whereas in the lower-diagonal, if the title is `x & y`, you get a plot of `corr[x(t+Lag), y(t)]` (yes, it is the same thing, but the lags are negative in the lower diagonal). The graphic is labeled in a strange way, just remember the second named series is the one that leads. In Figure 5.13 we notice that most of the correlations in the residual series are negligible, however, the zero-order correlations of mortality with temperature residuals is about .22 and mortality with particulate residuals is about .28 (type `acf(resid(fit),52)$acf`) to see the actual values. This means that the AR model is not capturing the concurrent effect of temperature and pollution on mortality (recall the data evolves over a week). It is possible to fit simultaneous models; see Reinsel (1997) for further details. Thus, not unexpectedly, the Q-test rejects the null hypothesis that the noise is white. The Q-test statistic is given by

$$Q = n^2 \sum_{h=1}^{H} \frac{1}{n-h} \operatorname{tr}\left[\widehat{\Gamma}_w(h)\widehat{\Gamma}_w(0)^{-1}\widehat{\Gamma}_w(h)\widehat{\Gamma}_w(0)^{-1}\right], \tag{5.91}$$

where

$$\widehat{\Gamma}_w(h) = n^{-1} \sum_{t=1}^{n-h} \widehat{\boldsymbol{w}}_{t+h}\widehat{\boldsymbol{w}}_t',$$

and $\widehat{\boldsymbol{w}}_t$ is the residual process. Under the null that $\boldsymbol{w}_t$ is white noise, (5.91) has an asymptotic χ^2 distribution with $k^2(H-p)$ degrees of freedom.

Finally, prediction follows in a straight forward manner from the univariate case. Using the R package `vars`, use the `predict` command and the `fanchart` command, which produces a nice graphic:

```
(fit.pr = predict(fit, n.ahead = 24, ci = 0.95))   # 4 weeks ahead
fanchart(fit.pr)   # plot prediction + error
```

The results are displayed in Figure 5.14; we note that the package stripped time when plotting the fanchart and the horizontal axis is labeled $1, 2, 3, \ldots$.

For pure VAR(p) models, the autocovariance structure leads to the multivariate version of the Yule–Walker equations:

$$\Gamma(h) = \sum_{j=1}^{p} \Phi_j \Gamma(h-j), \quad h = 1, 2, ..., \tag{5.92}$$

$$\Gamma(0) = \sum_{j=1}^{p} \Phi_j \Gamma(-j) + \Sigma_w. \tag{5.93}$$

where $\Gamma(h) = \operatorname{cov}(\boldsymbol{x}_{t+h}, \boldsymbol{x}_t)$ is a $k \times k$ matrix and $\Gamma(-h) = \Gamma(h)'$.

Estimation of the autocovariance matrix is similar to the univariate case, that is, with $\bar{\boldsymbol{x}} = n^{-1}\sum_{t=1}^{n} \boldsymbol{x}_t$, as an estimate of $\boldsymbol{\mu} = E\boldsymbol{x}_t$,

$$\widehat{\Gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (\boldsymbol{x}_{t+h} - \bar{\boldsymbol{x}})(\boldsymbol{x}_t - \bar{\boldsymbol{x}})', \quad h = 0, 1, 2, .., n-1, \tag{5.94}$$

and $\widehat{\Gamma}(-h) = \widehat{\Gamma}(h)'$. If $\widehat{\gamma}_{i,j}(h)$ denotes the element in the i-th row and j-th column of $\widehat{\Gamma}(h)$, the cross-correlation functions (CCF), as discussed in (1.35), are estimated by

$$\widehat{\rho}_{i,j}(h) = \frac{\widehat{\gamma}_{i,j}(h)}{\sqrt{\widehat{\gamma}_{i,i}(0)}\sqrt{\widehat{\gamma}_{j,j}(0)}} \quad h = 0, 1, 2, .., n-1. \tag{5.95}$$

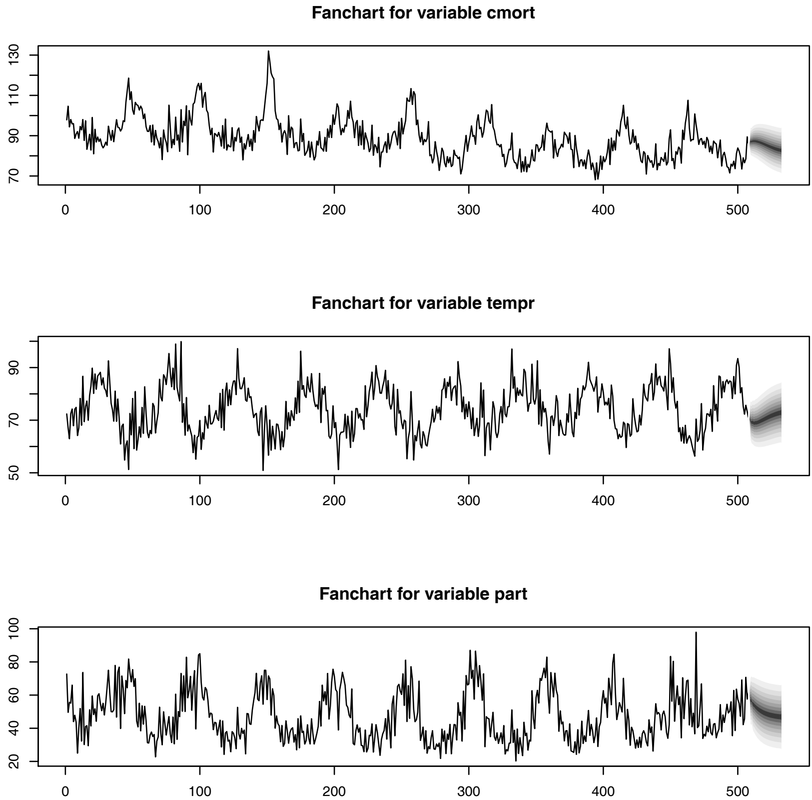

Fig. 5.14. Predictions from a VAR(2) fit to the LA mortality – pollution data.

When $i = j$ in (5.95), we get the estimated autocorrelation function (ACF) of the individual series.

Although least squares estimation was used in Examples 5.10 and 5.11, we could have also used Yule-Walker estimation, conditional or unconditional maximum likelihood estimation. As in the univariate case, the Yule–Walker estimators, the maximum likelihood estimators, and the least squares estimators are asymptotically equivalent. To exhibit the asymptotic distribution of the autoregression parameter estimators, we write

$$\boldsymbol{\phi} = \operatorname{vec}(\Phi_1, ..., \Phi_p),$$

where the vec operator stacks the columns of a matrix into a vector. For example, for a bivariate AR(2) model,

$$\boldsymbol{\phi} = \operatorname{vec}(\Phi_1, \Phi_2) = (\Phi_{1_{11}}, \Phi_{1_{21}}, \Phi_{1_{12}}, \Phi_{1_{22}} \Phi_{2_{11}}, \Phi_{2_{21}}, \Phi_{2_{12}}, \Phi_{2_{22}})',$$

where $\Phi_{\ell_{ij}}$ is the ij-th element of Φ_ℓ, $\ell = 1, 2$. Because $(\Phi_1, ..., \Phi_p)$ is a $k \times kp$ matrix, $\boldsymbol{\phi}$ is a $k^2 p \times 1$ vector. We now state the following property.

Property 5.1 Large-Sample Distribution of the Vector Autoregression Estimators

Let $\widehat{\boldsymbol{\phi}}$ denote the vector of parameter estimators (obtained via Yule–Walker, least squares, or maximum likelihood) for a k-dimensional AR(p) model. Then,

$$\sqrt{n}\left(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \sim AN(\mathbf{0}, \Sigma_w \otimes \Gamma_{pp}^{-1}), \tag{5.96}$$

where $\Gamma_{pp} = \{\Gamma(i-j)\}_{i,j=1}^{p}$ is a $kp \times kp$ matrix and $\Sigma_w \otimes \Gamma_{pp}^{-1} = \{\sigma_{ij}\Gamma_{pp}^{-1}\}_{i,j=1}^{k}$ is a $k^2p \times k^2p$ matrix with σ_{ij} denoting the ij-th element of Σ_w.

The variance–covariance matrix of the estimator $\widehat{\boldsymbol{\phi}}$ is approximated by replacing Σ_w by $\widehat{\Sigma}_w$, and replacing $\Gamma(h)$ by $\widehat{\Gamma}(h)$ in Γ_{pp}. The square root of the diagonal elements of $\widehat{\Sigma}_w \otimes \widehat{\Gamma}_{pp}^{-1}$ divided by $\sqrt{n}$ gives the individual standard errors. For the mortality data example, the estimated standard errors for the VAR(2) fit are listed in Example 5.11; although those standard errors were taken from a regression run, they could have also been calculated using Property 5.1.

A $k \times 1$ vector-valued time series $\boldsymbol{x}_t$, for $t = 0, \pm1, \pm2, \ldots$, is said to be VARMA$(p, q)$ if $\boldsymbol{x}_t$ is stationary and

$$\boldsymbol{x}_t = \boldsymbol{\alpha} + \Phi_1\boldsymbol{x}_{t-1} + \cdots + \Phi_p\boldsymbol{x}_{t-p} + \boldsymbol{w}_t + \Theta_1\boldsymbol{w}_{t-1} + \cdots + \Theta_q\boldsymbol{w}_{t-q}, \tag{5.97}$$

with $\Phi_p \neq 0$, $\Theta_q \neq 0$, and $\Sigma_w > 0$ (that is, Σ_w is positive definite). The coefficient matrices Φ_j; $j = 1, ..., p$ and Θ_j; $j = 1, ..., q$ are, of course, $k \times k$ matrices. If $\boldsymbol{x}_t$ has mean $\boldsymbol{\mu}$ then $\boldsymbol{\alpha} = (I - \Phi_1 - \cdots - \Phi_p)\boldsymbol{\mu}$. As in the univariate case, we will have to place a number of conditions on the multivariate ARMA model to ensure the model is unique and has desirable properties such as causality. These conditions will be discussed shortly.

As in the VAR model, the special form assumed for the constant component can be generalized to include a fixed $r \times 1$ vector of inputs, $\boldsymbol{u}_t$. That is, we could have proposed the vector ARMAX model,

$$\boldsymbol{x}_t = \Gamma\boldsymbol{u}_t + \sum_{j=1}^{p}\Phi_j\boldsymbol{x}_{t-j} + \sum_{k=1}^{q}\Theta_k\boldsymbol{w}_{t-k} + \boldsymbol{w}_t, \tag{5.98}$$

where Γ is a $p \times r$ parameter matrix.

While extending univariate AR (or pure MA) models to the vector case is fairly easy, extending univariate ARMA models to the multivariate case is not a simple matter. Our discussion will be brief, but interested readers can get more details in Lütkepohl (1993), Reinsel (1997), and Tiao and Tsay (1989).

In the multivariate case, the autoregressive operator is

$$\Phi(B) = I - \Phi_1 B - \cdots - \Phi_p B^p, \tag{5.99}$$

and the moving average operator is

$$\Theta(B) = I + \Theta_1 B + \cdots + \Theta_q B^q, \tag{5.100}$$

The zero-mean VARMA(p, q) model is then written in the concise form as

$$\Phi(B)\boldsymbol{x}_t = \Theta(B)\boldsymbol{w}_t. \tag{5.101}$$

The model is said to be causal if the roots of $|\Phi(z)|$ (where $|\cdot|$ denotes determinant) are outside the unit circle, $|z| > 1$; that is, $|\Phi(z)| \neq 0$ for any value z such that $|z| \leq 1$. In this case, we can write

$$\boldsymbol{x}_t = \Psi(B)\boldsymbol{w}_t,$$

where $\Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j$, $\Psi_0 = I$, and $\sum_{j=0}^{\infty} ||\Psi_j|| < \infty$. The model is said to be invertible if the roots of $|\Theta(z)|$ lie outside the unit circle. Then, we can write

$$\boldsymbol{w}_t = \Pi(B)\boldsymbol{x}_t,$$

where $\Pi(B) = \sum_{j=0}^{\infty} \Pi_j B^j$, $\Pi_0 = I$, and $\sum_{j=0}^{\infty} ||\Pi_j|| < \infty$. Analogous to the univariate case, we can determine the matrices Ψ_j by solving $\Psi(z) = \Phi(z)^{-1}\Theta(z)$, $|z| \leq 1$, and the matrices Π_j by solving $\Pi(z) = \Theta(z)^{-1}\Phi(z)$, $|z| \leq 1$.

For a causal model, we can write $\boldsymbol{x}_t = \Psi(B)\boldsymbol{w}_t$ so the general autocovariance structure of an ARMA(p, q) model is

$$\Gamma(h) = \text{cov}(\boldsymbol{x}_{t+h}, \boldsymbol{x}_t) = \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma_w \Psi_j'. \tag{5.102}$$

and $\Gamma(-h) = \Gamma(h)'$. For pure MA(q) processes, (5.102) becomes

$$\Gamma(h) = \sum_{j=0}^{q-h} \Theta_{j+h} \Sigma_w \Theta_j', \tag{5.103}$$

where $\Theta_0 = I$. Of course, (5.103) implies $\Gamma(h) = 0$ for $h > q$.

As in the univariate case, we will need conditions for model uniqueness. These conditions are similar to the condition in the univariate case that the autoregressive and moving average polynomials have no common factors. To explore the uniqueness problems that we encounter with multivariate ARMA models, consider a bivariate AR(1) process, $\boldsymbol{x}_t = (x_{t,1}, x_{t,2})'$, given by

$$\begin{aligned} x_{t,1} &= \phi x_{t-1,2} + w_{t,1}, \\ x_{t,2} &= w_{t,2}, \end{aligned}$$

where $w_{t,1}$ and $w_{t,2}$ are independent white noise processes and $|\phi| < 1$. Both processes, $x_{t,1}$ and $x_{t,2}$ are causal and invertible. Moreover, the processes are jointly stationary because $\text{cov}(x_{t+h,1}, x_{t,2}) = \phi\, \text{cov}(x_{t+h-1,2}, x_{t,2}) \equiv \phi\, \gamma_{2,2}(h-1) = \phi\sigma_{w_2}^2 \delta_1^h$ does not depend on t; note, $\delta_1^h = 1$ when $h = 1$, otherwise, $\delta_1^h = 0$. In matrix notation, we can write this model as

$$\boldsymbol{x}_t = \Phi \boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \quad \text{where} \quad \Phi = \begin{bmatrix} 0 & \phi \\ 0 & 0 \end{bmatrix}. \tag{5.104}$$

We can write (5.104) in operator notation as

$$\Phi(B)\boldsymbol{x}_t = \boldsymbol{w}_t \quad \text{where} \quad \Phi(z) = \begin{bmatrix} 1 & -\phi z \\ 0 & 1 \end{bmatrix}.$$

In addition, model (5.104) can be written as a bivariate ARMA(1,1) model

$$\boldsymbol{x}_t = \Phi_1 \boldsymbol{x}_{t-1} + \Theta_1 \boldsymbol{w}_{t-1} + \boldsymbol{w}_t, \tag{5.105}$$

where

$$\Phi_1 = \begin{bmatrix} 0 & \phi + \theta \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \Theta_1 = \begin{bmatrix} 0 & -\theta \\ 0 & 0 \end{bmatrix},$$

and θ is arbitrary. To verify this, we write (5.105), as $\Phi_1(B)\boldsymbol{x}_t = \Theta_1(B)\boldsymbol{w}_t$, or

$$\Theta_1(B)^{-1}\Phi_1(B)\boldsymbol{x}_t = \boldsymbol{w}_t,$$

where

$$\Phi_1(z) = \begin{bmatrix} 1 & -(\phi + \theta)z \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \Theta_1(z) = \begin{bmatrix} 1 & -\theta z \\ 0 & 1 \end{bmatrix}.$$

Then,

$$\Theta_1(z)^{-1}\Phi_1(z) = \begin{bmatrix} 1 & \theta z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -(\phi + \theta)z \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\phi z \\ 0 & 1 \end{bmatrix} = \Phi(z),$$

where $\Phi(z)$ is the polynomial associated with the bivariate AR(1) model in (5.104). Because θ is arbitrary, the parameters of the ARMA(1,1) model given in (5.105) are not identifiable. No problem exists, however, in fitting the AR(1) model given in (5.104).

The problem in the previous discussion was caused by the fact that both $\Theta(B)$ and $\Theta(B)^{-1}$ are finite; such a matrix operator is called unimodular. If $U(B)$ is unimodular, $|U(z)|$ is constant. It is also possible for two seemingly different multivariate ARMA(p, q) models, say, $\Phi(B)\boldsymbol{x}_t = \Theta(B)\boldsymbol{w}_t$ and $\Phi_*(B)\boldsymbol{x}_t = \Theta_*(B)\boldsymbol{w}_t$, to be related through a unimodular operator, $U(B)$ as $\Phi_*(B) = U(B)\Phi(B)$ and $\Theta_*(B) = U(B)\Theta(B)$, in such a way that the orders of $\Phi(B)$ and $\Theta(B)$ are the same as the orders of $\Phi_*(B)$ and $\Theta_*(B)$, respectively. For example, consider the bivariate ARMA(1,1) models given by

$$\Phi \boldsymbol{x}_t \equiv \begin{bmatrix} 1 & -\phi B \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_t = \begin{bmatrix} 1 & \theta B \\ 0 & 1 \end{bmatrix} \boldsymbol{w}_t \equiv \Theta w_t$$

and

$$\Phi_*(B)\boldsymbol{x}_t \equiv \begin{bmatrix} 1 & (\alpha - \phi)B \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_t = \begin{bmatrix} 1 & (\alpha + \theta)B \\ 0 & 1 \end{bmatrix} \boldsymbol{w}_t \equiv \Theta_*(B)\boldsymbol{w}_t,$$

where α, ϕ, and θ are arbitrary constants. Note,

$$\Phi_*(B) \equiv \begin{bmatrix} 1 & (\alpha-\phi)B \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \alpha B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\phi B \\ 0 & 1 \end{bmatrix} \equiv U(B)\Phi(B)$$

and

$$\Theta_*(B) \equiv \begin{bmatrix} 1 & (\alpha+\theta)B \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \alpha B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \theta B \\ 0 & 1 \end{bmatrix} \equiv U(B)\Theta(B).$$

In this case, both models have the same infinite MA representation $\boldsymbol{x}_t = \Psi(B)\boldsymbol{w}_t$, where

$$\Psi(B) = \Phi(B)^{-1}\Theta(B) = \Phi(B)^{-1}U(B)^{-1}U(B)\Theta(B) = \Phi_*(B)^{-1}\Theta_*(B).$$

This result implies the two models have the same autocovariance function $\Gamma(h)$. Two such ARMA(p,q) models are said to be observationally equivalent.

As previously mentioned, in addition to requiring causality and invertiblity, we will need some additional assumptions in the multivariate case to make sure that the model is unique. To ensure the identifiability of the parameters of the multivariate ARMA(p,q) model, we need the following additional two conditions: (i) the matrix operators $\Phi(B)$ and $\Theta(B)$ have no common left factors other than unimodular ones [that is, if $\Phi(B) = U(B)\Phi_*(B)$ and $\Theta(B) = U(B)\Theta_*(B)$, the common factor must be unimodular] and (ii) with q as small as possible and p as small as possible for that q, the matrix $[\Phi_p, \Theta_q]$ must be full rank, k. One suggestion for avoiding most of the aforementioned problems is to fit only vector AR(p) models in multivariate situations. Although this suggestion might be reasonable for many situations, this philosophy is not in accordance with the law of parsimony because we might have to fit a large number of parameters to describe the dynamics of a process.

Asymptotic inference for the general case of vector ARMA models is more complicated than pure AR models; details can be found in Reinsel (1997) or Lütkepohl (1993), for example. We also note that estimation for VARMA models can be recast into the problem of estimation for state-space models that will be discussed in Chapter 6.

A simple algorithm for fitting vector ARMA models from Spliid (1983) is worth mentioning because it repeatedly uses the multivariate regression equations. Consider a general ARMA(p,q) model for a time series with a nonzero mean

$$\boldsymbol{x}_t = \boldsymbol{\alpha} + \Phi_1\boldsymbol{x}_{t-1} + \cdots + \Phi_p\boldsymbol{x}_{t-p} + \boldsymbol{w}_t + \Theta_1\boldsymbol{w}_{t-1} + \cdots + \Theta_q\boldsymbol{w}_{t-q}. \tag{5.106}$$

If $\boldsymbol{\mu} = E\boldsymbol{x}_t$, then $\boldsymbol{\alpha} = (I - \Phi_1 - \cdots - \Phi_p)\boldsymbol{\mu}$. If $\boldsymbol{w}_{t-1}, ..., \boldsymbol{w}_{t-q}$ were observed, we could rearrange (5.106) as a multivariate regression model

$$\boldsymbol{x}_t = \mathcal{B}\boldsymbol{z}_t + \boldsymbol{w}_t, \tag{5.107}$$

with