THE UNIVERSITY OF CHICAGO
Booth School of Business
Business 41912, Spring Quarter 2020, Mr. Ruey S. Tsay

Lecture 5: Multivariate Multiple Linear Regression

The model is

$$\boldsymbol{Y}_{n\times m} = \boldsymbol{Z}_{n\times(r+1)}\boldsymbol{\beta}_{(r+1)\times m} + \boldsymbol{\epsilon}_{n\times m}, \tag{1}$$

where $E(\boldsymbol{\epsilon}_{(i)}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_{(i)}, \boldsymbol{\epsilon}_{(j)}) = \sigma_{ij}\boldsymbol{I}$ for $i, j = 1, \ldots, m$ with $\epsilon_{(i)}$ being the ith column of $\boldsymbol{\epsilon}$. The design matrix $\boldsymbol{Z}$ is of full rank $r+1$ and $n > r+1$.
Here we have n observations in $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Each observation $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})'$ contains m variables. Each observation $\boldsymbol{Z}_i = (1, Z_{i1}, \ldots, Z_{ir})'$ contains r measurements of the design matrix. Let $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im})'$ be the error term for the i observation and we have $\text{Cov}(\boldsymbol{\epsilon}_i)$ $= \boldsymbol{\Sigma} = [\sigma_{jk}]$. Note that in Eq.(1), the ith row of $\boldsymbol{Y}$ is $\boldsymbol{Y}_i'$, the ith row of $\boldsymbol{\epsilon}$ is $\boldsymbol{\epsilon}_i'$ and for the ith data point, we have

$$\boldsymbol{Y}_i = \boldsymbol{\beta}'\boldsymbol{Z}_i + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{Z}_i'$ is the ith row of $\boldsymbol{Z}$. That is, $\boldsymbol{Z}_i$ is a $(r+1)$-dimensional vector. The prior equation states the equation for the ith data point. We can also consider the jth component of $\boldsymbol{Y}_i$ over all sample. Then, we have

$$\boldsymbol{Y}_j = \boldsymbol{Z}\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j,$$

where $\boldsymbol{Y}_j$, $\boldsymbol{\beta}_j$ and $\boldsymbol{\epsilon}_j$ are the jth column of $\boldsymbol{Y}$, $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$, respectively. The prior equation is a multiple linear regression discussed before.
From Eq.(1), we have

$$\text{vec}(\boldsymbol{Y}) = (\boldsymbol{I}_m \otimes \boldsymbol{Z})\text{vec}(\boldsymbol{\beta}) + \text{vec}(\boldsymbol{\epsilon}), \tag{2}$$

where $\text{Cov}(\text{vec}(\boldsymbol{\epsilon})) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$. Here we use the identity $\text{vec}(\boldsymbol{AB}) = (\boldsymbol{I} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$.
The generalized least squares estimate of $\boldsymbol{\beta}$ is obtained by minimizing the objective function

$$\begin{aligned} S(\boldsymbol{\beta}) &= [\text{vec}(\boldsymbol{\epsilon})]'[\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n]^{-1}\text{vec}(\boldsymbol{\epsilon}) \\ &= [\text{vec}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})]'[\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n]\text{vec}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}) \\ &= tr[(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})']. \end{aligned}$$

The last equality holds because $\boldsymbol{\Sigma}$ is symmetric and we use

$$tr(\boldsymbol{DBC}) = [\text{vec}(\boldsymbol{C}')]'(\boldsymbol{B}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{D}).$$

Next, using $\text{vec}(\boldsymbol{Z}\boldsymbol{\beta}) = (\boldsymbol{I}_m \otimes \boldsymbol{Z})\text{vec}(\boldsymbol{\beta})$ and Eq.(1), we have

$$\begin{aligned} S(\boldsymbol{\beta}) &= [\text{vec}(\boldsymbol{Y})' - \text{vec}(\boldsymbol{\beta})'(\boldsymbol{I}_m \otimes \boldsymbol{Z}')][\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n][\text{vec}(\boldsymbol{Y}) - (\boldsymbol{I}_m \otimes \boldsymbol{Z})\text{vec}(\boldsymbol{\beta}))] \\ &= \text{vec}(\boldsymbol{Y})'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\text{vec}(\boldsymbol{Y}) - 2\text{vec}(\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}')\text{vec}(\boldsymbol{Y}) \\ &+ \text{vec}(\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}'\boldsymbol{Z})\text{vec}(\boldsymbol{\beta}). \end{aligned}$$

Taking partial derivative of $S(\boldsymbol{\beta})$ with respect to $\text{vec}(\boldsymbol{\beta})$, we have

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \text{vec}(\boldsymbol{\beta})} = -2(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}')\text{vec}(\boldsymbol{Y}) + 2(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}'\boldsymbol{Z})\text{vec}(\boldsymbol{\beta}).$$

Equating to zero gives the normal equations

$$(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}'\boldsymbol{Z})\text{vec}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}')\text{vec}(\boldsymbol{Y}).$$

Consequently, the GLS estimate is

$$\begin{aligned}
\text{vec}(\widehat{\boldsymbol{\beta}}) &= (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}')\text{vec}(\boldsymbol{Y}) \\
&= [\boldsymbol{\Sigma} \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}](\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}')\text{vec}(\boldsymbol{Y}) \\
&= [\boldsymbol{I}_m \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}']\text{vec}(\boldsymbol{Y}) \qquad (3) \\
&= \text{vec}[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}].
\end{aligned}$$

Therefore,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{Y}).$$

Note that this is the ordinary least squares estimate. Also, from Eq. (3), we have

$$\begin{aligned}
\text{Cov}[\text{vec}(\widehat{\boldsymbol{\beta}})] &= [\boldsymbol{I}_m \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'][\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n][\boldsymbol{I}_m \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}']' \\
&= [\boldsymbol{I}_m \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'][\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n][\boldsymbol{I}_m \otimes \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}] \\
&= \boldsymbol{\Sigma} \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}.
\end{aligned}$$

In a sense, we put m multiple linear regressions with the same design matrix $\boldsymbol{Z}$ together. The errors between the ith and kth multiple linear regressions might be correated as $\sigma_{ik}\boldsymbol{I}_n$. Let the ith column of $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_{(i)}$ for $i = 1, \ldots, m$. We can think of the model as one that consists of m multiple lienar regressions with the same design matrix $\boldsymbol{Z}$. As such, we can estimate $\boldsymbol{\beta}_{(i)}$ as

$$\widehat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_{(i)}.$$

Collecting the estimates, we obtain

$$\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_{(1)}, \widehat{\boldsymbol{\beta}}_{(2)}, \ldots, \widehat{\boldsymbol{\beta}}_{(m)}] = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'[\boldsymbol{Y}_{(1)}, \boldsymbol{Y}_{(2)}, \ldots, \boldsymbol{Y}_{(m)}],$$

or, for simplicity,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}.$$

For any choice of parameters $\boldsymbol{B} = [\boldsymbol{B}_{(1)}, \boldsymbol{B}_{(2)}, \ldots, \boldsymbol{B}_{(m)}]$, the error matrix is $\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{B}$. The error sum of squares and cross-products matrix is

$$(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{B}) =$$

$$\begin{bmatrix} (\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{B}_{(1)})'(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{B}_{(1)}) & \cdots & (\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{B}_{(1)})'(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{B}_{(m)}) \\ \vdots & & \vdots \\ (\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{B}_{(m)})'(\boldsymbol{Y}_{(1)} - \boldsymbol{Z}\boldsymbol{B}_{(1)}) & \cdots & (\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{B}_{(m)})'(\boldsymbol{Y}_{(m)} - \boldsymbol{Z}\boldsymbol{B}_{(m)}) \end{bmatrix}.$$

The least squares estimates $\widehat{\boldsymbol{\beta}}_{(i)}$ thus minimize $tr[(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{B})]$. The estimates also minimize the generalized variance $|(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{Z}\boldsymbol{B})|$.
Let $\widehat{\boldsymbol{Y}} = \boldsymbol{Z}\widehat{\boldsymbol{\beta}}$ be the fitted values and $\widehat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ be the residuals. It follows that $\widehat{\boldsymbol{\epsilon}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$, where $\boldsymbol{H}$ is the hat-matrix as defined in the multiple linear regression. The orthogonality properties of the multiple linear regression continue to hold for the multivariate multiple linear regression. Specfically,

1. $\boldsymbol{Z}'\widehat{\boldsymbol{\epsilon}} = \mathbf{0}$,
2. $\widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{\epsilon}} = \mathbf{0}$,
3. $\boldsymbol{Y}'\boldsymbol{Y} = \widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{Y}} + \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}$, i.e. $\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = \boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{Z}'\boldsymbol{Z}\widehat{\boldsymbol{\beta}}$.

The second property says that $\widehat{\boldsymbol{Y}}_{(i)}$ is perpendicular to all residual vetors $\widehat{\boldsymbol{\epsilon}}_{(k)}$.

Result 7.9. For the LSE $\widehat{\boldsymbol{\beta}}$ with $\boldsymbol{Z}$ being of full rank $r+1$,

1. $E(\widehat{\boldsymbol{\beta}}_{(i)}) = \boldsymbol{\beta}_{(i)}$, i.e. $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
2. $\text{Cov}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$, $i, j = 1, \ldots, m$.
3. $E(\widehat{\boldsymbol{\epsilon}}) = \mathbf{0}$ and $E(\widehat{\boldsymbol{\epsilon}}_{(i)}, \widehat{\boldsymbol{\epsilon}}_{(k)}) = (n-r-1)\sigma_{ik}$, so $E(\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}) = (n-r-1)\boldsymbol{\Sigma}$.
4. $\widehat{\boldsymbol{\epsilon}}$ and $\widehat{\boldsymbol{\beta}}$ are uncorrelated.

Proof: Use the results of multiple linear regression discussed before. In addition,

$$\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_{(i)} - \boldsymbol{\beta}_{(i)} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\epsilon}_{(i)},$$

and

$$\widehat{\boldsymbol{\epsilon}}_{(i)} = \boldsymbol{Y}_{(i)} - \widehat{\boldsymbol{Y}}_{(i)} = [\boldsymbol{I} - \boldsymbol{H}]\boldsymbol{\epsilon}_{(i)}.$$

Result 7.10. Assume that $\boldsymbol{Z}$ is of full rank $r+1$ and $n \geq (r+1)+m$. If the errors $\boldsymbol{\epsilon}$ have a multivariate normal distribution, then the LSE $\widehat{\boldsymbol{\beta}}$ is also the maximum likelihood estimator of $\boldsymbol{\beta}$, and $\widehat{\boldsymbol{\beta}}$ has a normal distribution with mean $\boldsymbol{\beta}$ and $\text{Cov}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$. In addition, $\widehat{\boldsymbol{\beta}}$ is independent of the MLE of the positive definite matrix $\boldsymbol{\Sigma}$ given by

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}),$$

and $n\widehat{\boldsymbol{\Sigma}}$ is distributed as $W_{p,n-r-1}(\boldsymbol{\Sigma})$. The maximized likelihood $L(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = (2\pi)^{-mn/2}|\widehat{\boldsymbol{\sigma}}|^{-n/2}e^{-mn/2}$.

1 Inference

Partitioning $\boldsymbol{Z} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2]$, where $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are $n \times (q+1)$ and $n \times (r-q)$ matrix, respectively. The model becomes

$$\boldsymbol{Y} = \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.$$

Test $H_o : \boldsymbol{\beta}_2 = \mathbf{0}$ vs $H_a : \boldsymbol{\beta}_2 \neq \mathbf{0}$.
As before, consider the submodel

$$\boldsymbol{Y} = \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}.$$

The LSE of $\boldsymbol{\beta}_1$ is $\widehat{\boldsymbol{\beta}}_1 = (\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1'\boldsymbol{Y}$ and, under normality, the estimate of the residual covariance matrix is

$$\widehat{\boldsymbol{\Sigma}}_1 = n^{-1}(\boldsymbol{Y} - \boldsymbol{Z}_1\widehat{\boldsymbol{\beta}}_1)'(\boldsymbol{Y} - \boldsymbol{Z}_1\boldsymbol{\beta}_1).$$

The likelihood ratio is

$$\Lambda = \frac{\max_{\boldsymbol{\beta}_1, \boldsymbol{\Sigma}} L(\boldsymbol{\beta}_1, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\beta}, \boldsymbol{\Sigma}} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})} = \frac{L(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\Sigma}}_1)}{L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}})} = \left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_1|}\right)^{n/2}.$$

Result 7.11. Assume that $\boldsymbol{Z}$ is of full rank $r+1$ and $n > m+r+1$. Also, $\boldsymbol{\epsilon}$ are normally distributed. Under $H_o : \boldsymbol{\beta}_2 = \mathbf{0}$, $n\widehat{\boldsymbol{\Sigma}}$ is distributed as $W_{p,n-r-1}(\boldsymbol{\Sigma})$ independently of $n(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}})$ which, in turn, is distributed as $W_{p,r-q}(\boldsymbol{\Sigma})$. The likelihood ratio test of H_o is equivalent to rejecting H_o for large values of

$$-2\ln(\Lambda) = -n \ln\left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_1|}\right).$$

Furthermore, for large n, the modified statistic

$$-\left[n - r - 1 - \frac{1}{2}(m - r + q + 1)\right] \ln\left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_1|}\right)$$

has, to a close approximation, a chi-square distribution with $m(r-q)$ degrees of freedom.

Remarks: Some alternative test statistics have been proposed in the literature for testing $H_o : \boldsymbol{\beta}_2 = \mathbf{0}$. Let $\boldsymbol{E} = n\widehat{\boldsymbol{\Sigma}}$, $\boldsymbol{G} = n(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}})$ and $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_s$, where $s = \min(p, r-q)$, be the non-zero eigenvalues of $\boldsymbol{G}\boldsymbol{E}^{-1}$. Then,

- Wilk's lambda $= \prod_{i=1}^{s} \frac{1}{1+\eta_i} = \frac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{G}|}$,
- Pillai's trace $= \sum_{i=1}^{s} \frac{\eta_i}{1+\eta_i} = tr[\boldsymbol{G}(\boldsymbol{G}+\boldsymbol{E})^{-1}]$,
- Hotelling-Lawley trace $= \sum_{i=1}^{s} \eta_i = tr[\boldsymbol{G}\boldsymbol{E}^{-1}]$,
- Roy's greatest root $= \frac{\eta_1}{1+\eta_1}$.

Prediction: To predict the mean response at a fixed values $\boldsymbol{Z}_0$ of the predictor variables, we have

$$\boldsymbol{Z}_0'\widehat{\boldsymbol{\beta}} \sim N_m(\boldsymbol{Z}_0'\boldsymbol{\beta}, \boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0\boldsymbol{\Sigma}),$$

and

$$n\widehat{\boldsymbol{\Sigma}} \sim W_{n-r-1}(\boldsymbol{\Sigma}),$$

where the above two statistics are independent.
The $100(1-\alpha)\%$ confidence ellipsoid for $\boldsymbol{Z}_0'\boldsymbol{\beta}$ is

$$\boldsymbol{Z}_0'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'\left(\frac{n}{n-r-1}\widehat{\boldsymbol{\Sigma}}\right)^{-1}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'\boldsymbol{Z}_0$$

$$\leq \boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0\left[\left(\frac{m(n-r-1)}{n-r-m}\right)F_{m,n-r-m}(\alpha)\right],$$

where $F_{m,n-r-m}(\alpha)$ is the upper 100αth percentile of an F-distribution with m and $n-r-m$ degrees of freedom.
The $100(1-\alpha)\%$ simultaneous confidence intervals for $E(Y_i) = \boldsymbol{Z}_0\boldsymbol{\beta}_{(i)}$ are

$$\boldsymbol{Z}_0'\widehat{\boldsymbol{\beta}}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right)F_{m,n-r-m}(\alpha)}\sqrt{\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0\left(\frac{n}{n-r-1}\hat{\sigma}_{ii}\right)},$$

where $\boldsymbol{\beta}_{(i)}$ is the ith column of $\widehat{\boldsymbol{\beta}}$ and $\hat{\sigma}_{ii}$ is the (i,i)th element of $\widehat{\boldsymbol{\Sigma}}$.
Forecasting: To forecast the response $\boldsymbol{Y}_0 = \boldsymbol{Z}_0'\boldsymbol{\beta} + \boldsymbol{\epsilon}_0$ at $\boldsymbol{Z}_0$, we have

$$\boldsymbol{Y}_0 - \boldsymbol{Z}_0'\widehat{\boldsymbol{\beta}}' = \boldsymbol{Z}_0'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) + \boldsymbol{\epsilon}_0 \sim N_m[\boldsymbol{0}, (1+\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0)\boldsymbol{\Sigma}],$$

which is independent of $n\widehat{\boldsymbol{\Sigma}}$. Thus, the $100(1-\alpha)\%$ prediction ellipsoid for $\boldsymbol{Y}_0$ is

$$(\boldsymbol{Y}_0 - \boldsymbol{Z}_0'\widehat{\boldsymbol{\beta}})'\left(\frac{n}{n-r-1}\widehat{\boldsymbol{\Sigma}}\right)^{-1}(\boldsymbol{Y}_0 - \boldsymbol{Z}_0'\widehat{\boldsymbol{\beta}})$$

$$\leq (1+\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0)\left[\left(\frac{m(n-r-1)}{n-r-m}F_{m,n-r-m}(\alpha)\right)\right].$$

The $100(1-\alpha)\%$ simultaneous prediction intervals for the individual responses Y_{0i} are

$$\boldsymbol{Z}_0'\widehat{\boldsymbol{\beta}}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}F_{m,n-r-m}(\alpha)\right)}\sqrt{(1+\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0)\left(\frac{n}{n-r-1}\hat{\sigma}_{ii}\right)},$$

for $i = 1, \ldots, m$.

2 Regression models with time-series errors

See R demonstration, using the Natural Gas Data on Table 7.4 of the textbook. In particular, a multiplicative seasonal model seems to fit the data better than the one used in the text.

Example 7.6 of the textbook: natural gas for heating. Daily data and the variables are

1. Y: (Dependent variable) sendouts of natural gas
2. X_1: Degree of heating days (DHD) which is defined as 65^o- daily average temperature.
3. X_2: Lagged DHD, i.e. DHD_{Lag}.
4. X_3: Windspeed (24-hour average)
5. X_4: Weekend indicator.

There are 63 observations shown in Table 7.4.