Estimation of Covariance Matrix

Zongwu Cai

April 4, 2023

1 Methodology

Consider a regression model stated in (1) below. There may exist situations which the error e_t has **serial correlations** and/or **conditional heteroscedasticity**, but the main objective of the analysis is to make inference concerning the regression coefficients $\boldsymbol{\beta}$. When e_t has serial correlations, we assume that e_t follows an ARIMA type model but this assumption might not be always satisfied in some applications. Here, we consider a general situation without making this assumption. In situations under which the ordinary least squares estimates of the coefficients remain consistent, methods are available to provide **consistent estimate of the covariance matrix** of the coefficients. Two such methods are widely used in economics and finance. The first method is called **heteroscedasticity consistent (HC) estimator**; see Eicker (1967) and White (1980). The second method is called **heteroscedasticity and autocorrelation consistent (HAC) estimator**; see Newey and West (1987).

To ease in discussion, we write a regression model as

$$y_t = \boldsymbol{\beta}^T \mathbf{x}_t + e_t, \tag{1}$$

where y_t is the dependent variable, $\mathbf{x}_t = (x_{1t}, \ \cdots, \ x_{pt})^T$ is a p-dimensional vector of explanatory variables including constant and lagged variables, and $\boldsymbol{\beta} = (\beta_1, \ \cdots, \ \beta_p)^T$ is the parameter vector. The LS estimate of $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{t=1}^{n} \mathbf{x}_t \, \mathbf{x}_t^T\right]^{-1} \sum_{t=1}^{n} \mathbf{x}_t \, y_t,$$

and the associated covariance matrix has the so-called "sandwich" form as

$$\Sigma_\beta = \text{Cov}(\widehat{\boldsymbol{\beta}}) = \left[\sum_{t=1}^{n} \mathbf{x}_t \, \mathbf{x}_t^T\right]^{-1} \mathbf{C} \left[\sum_{t=1}^{n} \mathbf{x}_t \, \mathbf{x}_t^T\right]^{-1} \overset{if\ e_t\ is\ iid}{=} \sigma_e^2 \left[\sum_{t=1}^{n} \mathbf{x}_t \, \mathbf{x}_t^T\right]^{-1},$$

where $\mathbf{C}$ is called the "meat" given by

$$\mathbf{C} = \mathrm{Var}\left(\sum_{t=1}^{n} e_t\,\mathbf{x}_t\right),$$

σ_e^2 is the variance of e_t and is estimated by the variance of residuals of the regression. In the presence of serial correlations or conditional heteroscedasticity, the prior covariance matrix estimator is inconsistent, often resulting in inflating the t-ratios of $\widehat{\boldsymbol{\beta}}$.

The estimator of White (1980) is based on following:

$$\widehat{\Sigma}_{\beta,hc} = \left[\sum_{t=1}^{n} \mathbf{x}_t\,\mathbf{x}_t^T\right]^{-1} \widehat{\mathbf{C}}_{hc} \left[\sum_{t=1}^{n} \mathbf{x}_t\,\mathbf{x}_t^T\right]^{-1},$$

where with $\widehat{e}_t = y_t - \widehat{\boldsymbol{\beta}}^T\,\mathbf{x}_t$ being the residual at time t,

$$\widehat{\mathbf{C}}_{hc} = \frac{n}{n-p}\sum_{t=1}^{n} \widehat{e}_t^2\,\mathbf{x}_t\,\mathbf{x}_t^T.$$

The estimator of Newey and West (1987) is

$$\widehat{\Sigma}_{\beta,hac} = \left[\sum_{t=1}^{n} \mathbf{x}_t\,\mathbf{x}_t^T\right]^{-1} \widehat{\mathbf{C}}_{hac} \left[\sum_{t=1}^{n} \mathbf{x}_t\,\mathbf{x}_t^T\right]^{-1},$$

where $\widehat{\mathbf{C}}_{hac}$ is given by

$$\widehat{\mathbf{C}}_{hac} = \sum_{t=1}^{n} \widehat{e}_t^2\,\mathbf{x}_t\,\mathbf{x}_t^T + \sum_{j=1}^{l} w_j \sum_{t=j+1}^{n} \left\{\mathbf{x}_t\,\widehat{e}_t\,\widehat{e}_{t-j}\,\mathbf{x}_{t-j}^T + \mathbf{x}_{t-j}\,\widehat{e}_{t-j}\,\widehat{e}_t\,\mathbf{x}_t^T\right\}$$

with l is a truncation parameter and w_j is weight function such as the Barlett weight function defined by $w_j = 1 - j/(l+1)$. Other weight function can also used. Newey and West (1987) showed that if $l \to \infty$ and $l^4/n \to 0$, then $\widehat{\mathbf{C}}_{hac}$ is a consistent estimator of C. Newey and West (1987) suggested choosing l to be the integer part of $4(n/100)^{1/4}$ and Newey and West (1994) suggested using some adaptive (data-driven) methods to choose l; see Newey and West (1994) for details. In general, this estimator essentially can use a nonparametric method to estimate the covariance matrix of $\sum_{t=1}^{n} e_t\,\mathbf{x}_t$ and a class of kernel-based **heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators** was introduced by Andrews (1991). For example, the Barlett weight w_j above can be replaced by $w_j = K(j/(l+1))$ where $K(\cdot)$ is a kernel function such as truncated kernel $K(x) = I(|x| \le 1)$, the Tukey-Hanning kernel $K(x) = (1+\cos(\pi\,x))/2$ if $|x| \le 1$, the Parzen kernel

$$K(x) = \begin{cases} 1 - 6\,x^2 + 6\,|x|^3 & \text{for } 0 \le |x| \le 1/2, \\ 2(1-|x|)^3 & \text{for } 1/2 \le |x| \le 1, \\ 0 & \text{otherwsie}, \end{cases}$$

and the Quadratic spectral kernel

$$K(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi\, x/5)}{6\pi\, x/5} - \cos(6\pi\, x/5) \right).$$

Andrews (1991) suggested using the data-driven method to select the bandwidth l: $\widehat{l} = 2.66(\widehat{\alpha}\, T)^{1/5}$ for the Parzen kernel, $\widehat{l} = 1.7462(\widehat{\alpha}\, T)^{1/5}$ for the Tukey-Hanning kernel, and $\widehat{l} = 1.3221(\widehat{\alpha}\, T)^{1/5}$ for the quadratic spectral kernel, where

$$\widehat{\alpha} = \frac{4 \sum_{i=1}^{k} \widehat{\rho}_i^2 \widehat{\sigma}_i^4 / (1 - \widehat{\rho}_i)^8}{\sum_{i=1}^{n} \widehat{\sigma}_i^4 / (1 - \widehat{\rho}_i)^4}$$

with $\widehat{\rho}_i$ and $\widehat{\sigma}_i$ being parameters estimated from an AR(1) model for $\widehat{u}_t = \widehat{e}_t\, \mathbf{x}_t$.

2 A Real Example

Example 1: We consider the relationship between two U.S. weekly interest rate series: x_t: the 1-year Treasury constant maturity rate and y_t: the 3-year Treasury constant maturity rate. Both series have 1967 observations from January 5, 1962 to September 10, 1999[1] and are measured in percentages. The series are obtained from the Federal Reserve Bank of St Louis.

Figure 1 shows the time plots of the two interest rates with solid line denoting the 1-year rate and dashed line for the 3-year rate. The left panel of Figure 2 plots y_t versus x_t, indicating that, as expected, the two interest rates are highly correlated. A naive way to describe the relationship between the two interest rates is to use the simple model, **Model I**: $y_t = \beta_1 + \beta_2\, x_t + e_t$. This results in a fitted model $y_t = 0.911 + 0.924\, x_t + e_t$, with $\widehat{\sigma}_e^2 = 0.538$ and $R^2 = 95.8\%$, where the standard errors of the two coefficients are 0.032 and 0.004, respectively. This simple model (Model I) confirms the high correlation between the two interest rates. However, the model is seriously inadequate as shown by Figure 3, which gives the time plot and ACF of its residuals. In particular, the sample ACF of the residuals is highly significant and decays slowly, showing the pattern of a **unit root** nonstationary time series. The behavior of the residuals suggests that marked differences exist between the two interest rates. Using the modern econometric terminology, if one assumes that the two interest rate series are unit root nonstationary, then the behavior of the residuals indicates that the two interest rates are not **co-integrated**. In other words, the data fail to support the hypothesis that there exists a long-term equilibrium between the two interest rates. In some sense, this is not surprising because the pattern of "inverted yield curve" did occur

[1]It would be better to update the data until now.

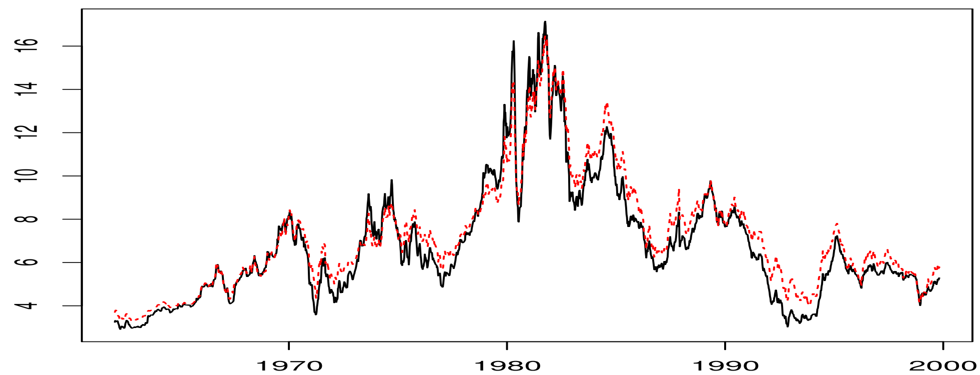

Figure 1: Time plots of U.S. weekly interest rates (in percentages) from January 5, 1962 to September 10, 1999. The solid line (black) is the Treasury 1-year constant maturity rate and the dashed line the Treasury 3-year constant maturity rate (red).

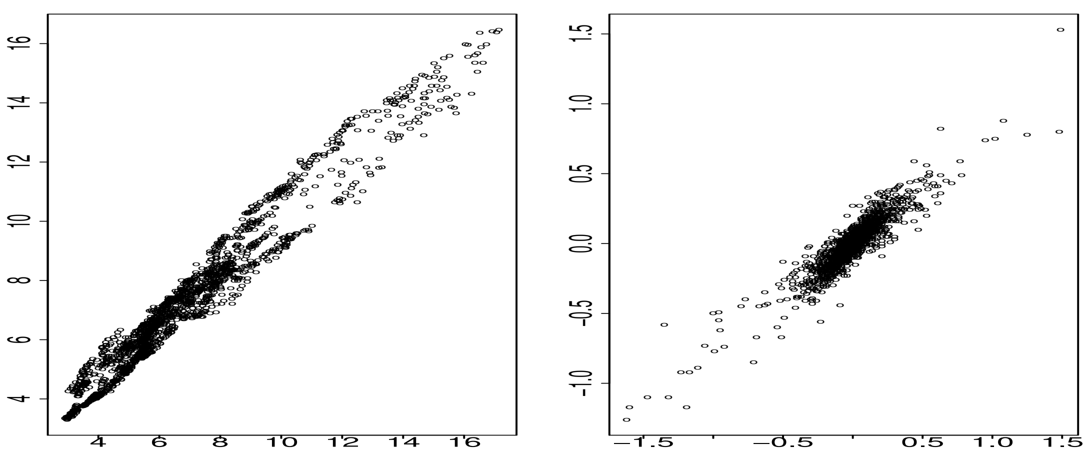

Figure 2: Scatterplots of U.S. weekly interest rates from January 5, 1962 to September 10, 1999: the left panel is 3-year rate versus 1-year rate, and the right panel is changes in 3-year rate versus changes in 1-year rate.

during the data span. By the inverted yield curve, we mean the situation under which interest rates are inversely related to their time to maturities.

The unit root behavior of both interest rates and the residuals leads to the consideration of the change series of interest rates. Let $\Delta x_t = y_t - y_{t-1} = (1-L)\,x_t$ be changes in the 1-year interest rate and $\Delta\,y_t = y_t - y_{t-1} = (1-L)\,y_t$ denote changes in the 3-year interest rate. Consider the linear regression, **Model II**: $\Delta\,y_t = \beta_1 + \beta_2\,\Delta\,x_t + e_t$. Figure 4 shows time plots of the two change series, whereas the right panel of Figure 2 provides a scatterplot between them. The change series remain highly correlated with a fitted linear regression

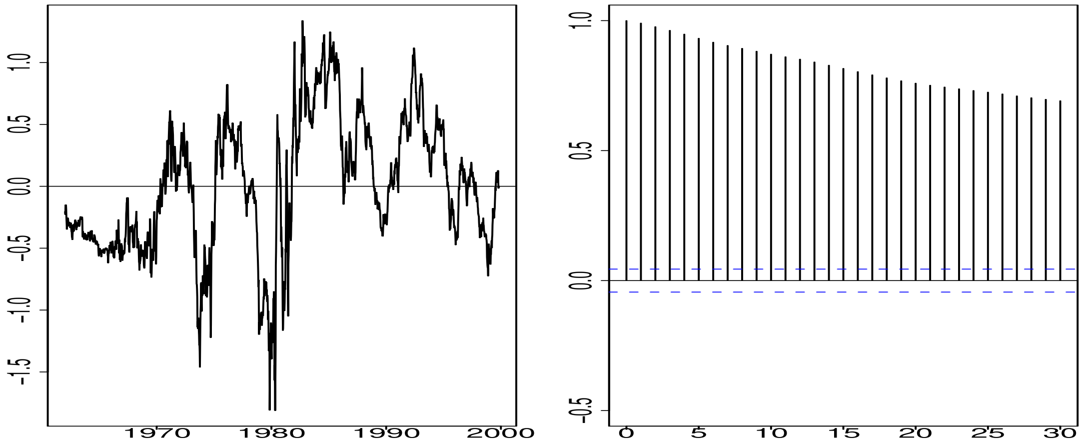

Figure 3: Residual series of linear regression Model I for two U.S. weekly interest rates: the left panel is time plot and the right panel is ACF.

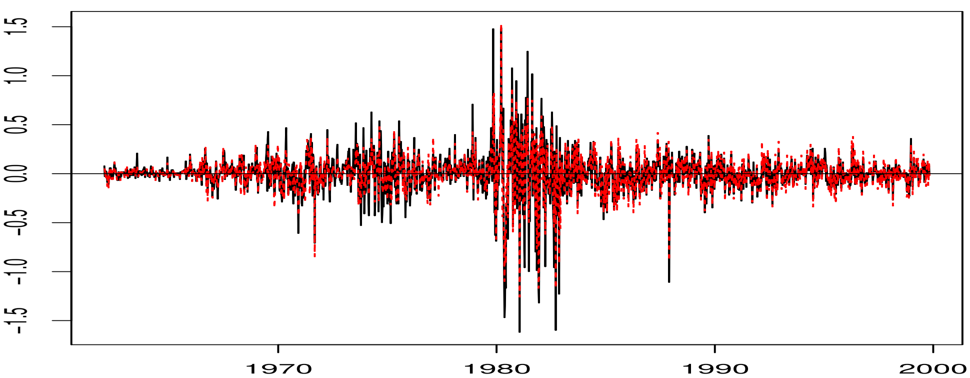

Figure 4: Time plots of the change series of U.S. weekly interest rates from January 12, 1962 to September 10, 1999: changes in the Treasury 1-year constant maturity rate are in denoted by black solid line, and changes in the Treasury 3-year constant maturity rate are indicated by red dashed line.

model given by $\Delta\, y_t = 0.0002 + 0.7811\,\Delta\, x_t + e_t$ with $\widehat{\sigma}_e^2 = 0.0682$ and $R^2 = 84.8\%$. The standard errors of the two coefficients are 0.0015 and 0.0075, respectively. This model further confirms the strong linear dependence between interest rates. The two top panels of Figure 5 show the time plot (left) and sample ACF (right) of the residuals (Model II). Once again, the ACF shows some significant serial correlation in the residuals, but the magnitude of the correlation is much smaller. This weak serial dependence in the residuals can be modeled by using the simple time series models discussed in the previous sections, and we have a linear regression with time series errors.

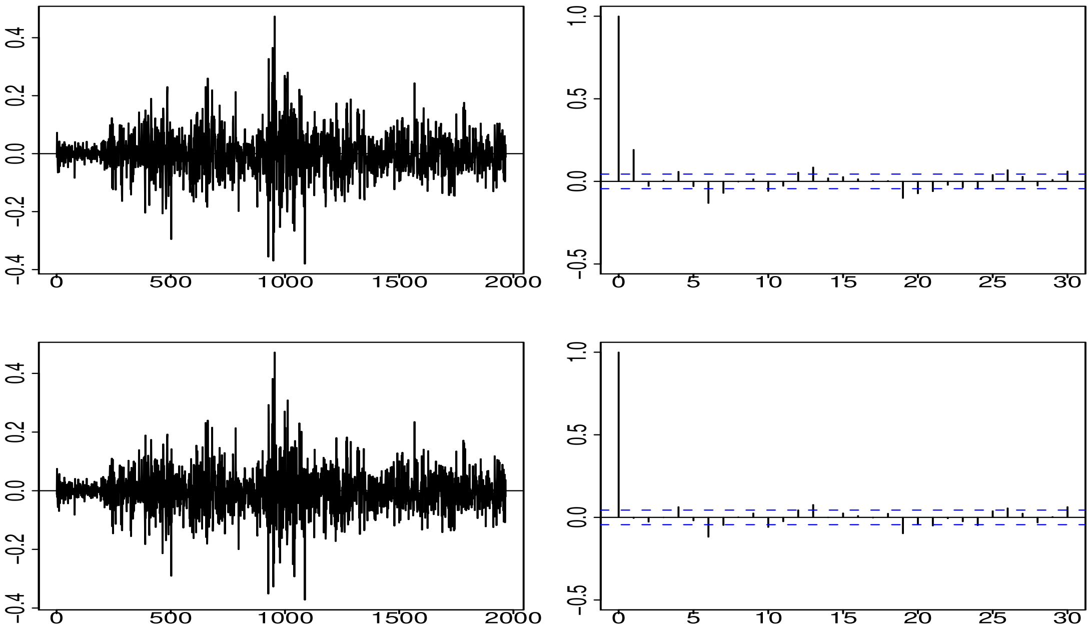

Figure 5: Residual series of the linear regression models: Model II (top) and Model III (bottom) for two change series of U.S. weekly interest rates: time plot (left) and ACF (right).

For illustration, we consider the first differenced interest rate series in Model II. The t-ratio of the coefficient of Δx_t is 104.63 if both serial correlation and conditional heteroscedasticity in residuals are ignored; it becomes 46.73 when the HC estimator is used, and it reduces to 40.08, when the HAC estimator is employed.

3 R Commands

To use HC or HAC estimator, we can use the package **sandwich** in **R** and the commands are **vcovHC()** or **vcovHAC()** or **meatHAC()**. There are a set of functions implementing a class of kernel-based heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators as introduced by Andrews (1991). In **vcovHC()**, these estimators differ in their choice of the ω_i in $\Omega = \text{Var}(e) = \text{diag}\{\omega_1, \cdots, \omega_n\}$, an overview of the most important cases is given in the following:

$$\begin{aligned} \text{const} : \omega_i &= \widehat{\sigma}^2 \\ HC0 : \omega_i &= \widehat{e}_i^2 \\ HC1 : \omega_i &= \frac{n}{n-k}\,\widehat{e}_i^2 \end{aligned}$$

$$
\begin{aligned}
HC2 : \omega_i &= \frac{\widehat{e}_i^2}{1-h_i} \\
HC3 : \omega_i &= \frac{\widehat{e}_i^2}{(1-h_i)^2} \\
HC4 : \omega_i &= \frac{\widehat{e}_i^2}{(1-h_i)^{\delta_i}}
\end{aligned}
$$

where $h_i = H_{ii}$ are the diagonal elements of the hat matrix and $\delta_i = \min\{4, h_i/\overline{h}\}$. For $HC4m$ and $HC5$, please see the paper by Cribari-Neto and da Silva (2011).

```
vcovHC(x, type = c("HC3", "const", "HC", "HC0", "HC1", "HC2", "HC4", "HC4m", "HC5"),
      omega = NULL, sandwich = TRUE, ...)

meatHC(x, type = , omega = NULL)

vcovHAC(x, order.by = NULL, prewhite = FALSE, weights = weightsAndrews,
      adjust = TRUE, diagnostics = FALSE, sandwich = TRUE, ar.method = "ols",
      data = list(), ...)

meatHAC(x, order.by = NULL, prewhite = FALSE, weights = weightsAndrews,
      adjust = TRUE, diagnostics = FALSE, ar.method = "ols", data = list())

kernHAC(x, order.by = NULL, prewhite = 1, bw = bwAndrews,
 kernel = c("Quadratic Spectral", "Truncated", "Bartlett", "Parzen",
"Tukey-Hanning"), approx = c("AR(1)", "ARMA(1,1)"), adjust = TRUE,
 diagnostics = FALSE, sandwich = TRUE, ar.method = "ols", tol = 1e-7,
 data = list(), verbose = FALSE, ...)

weightsAndrews(x, order.by = NULL,bw = bwAndrews,
 kernel = c("Quadratic Spectral","Truncated","Bartlett","Parzen",
 "Tukey-Hanning"), prewhite = 1, ar.method = "ols", tol = 1e-7,
 data = list(), verbose = FALSE, ...)

bwAndrews(x,order.by=NULL,kernel=c("Quadratic Spectral", "Truncated",
 "Bartlett","Parzen","Tukey-Hanning"), approx=c("AR(1)", "ARMA(1,1)"),
 weights = NULL, prewhite = 1, ar.method = "ols", data = list(), ...)
```

Also, there are a set of functions implementing the Newey and West (1987, 1994) heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators.

```
NeweyWest(x, lag = NULL, order.by = NULL, prewhite = TRUE, adjust = FALSE,
 diagnostics = FALSE, sandwich = TRUE, ar.method = "ols", data = list(),
 verbose = FALSE)

bwNeweyWest(x, order.by = NULL, kernel = c("Bartlett", "Parzen",
 "Quadratic Spectral", "Truncated", "Tukey-Hanning"), weights = NULL,
 prewhite = 1, ar.method = "ols", data = list(), ...)
```

4 Reading Materials – the papers by Zeileis (2004, 2006)

5 Computer Codes

```
#####################################################
# This is Example 1 for weekly interest rate series
#####################################################

z<-read.table("c:/res-teach/xiada/teaching05-07/data/ex2-1.txt",header=F)
#  first column=one year Treasury constant maturity rate;
# second column=three year Treasury constant maturity rate;
#  third column=date

x=z[,1]
y=z[,2]
n=length(x)
u=seq(1962+1/52,by=1/52,length=n)
x_diff=diff(x)
y_diff=diff(y)
# Fit a simple regression model and examine the residuals
fit1=lm(y~x)             # Model 1
e1=fit1$resid

postscript(file="c:/res-teach/xiada/teaching05-07/figs/fig-2.1.eps",
horizontal=F,width=6,height=6)
matplot(u,cbind(x,y),type="l",lty=c(1,2),col=c(1,2),ylab="",xlab="")
```

```
dev.off()

postscript(file="c:/res-teach/xiada/teaching05-07/figs/fig-2.2.eps",
horizontal=F,width=6,height=6)
par(mfrow=c(1,2),mex=0.4,bg="light grey")
plot(x,y,type="p",pch="o",ylab="",xlab="",cex=0.5)
plot(x_diff,y_diff,type="p",pch="o",ylab="",xlab="",cex=0.5)
dev.off()

postscript(file="c:/res-teach/xiada/teaching05-07/figs/fig-2.3.eps",
horizontal=F,width=6,height=6)
par(mfrow=c(1,2),mex=0.4,bg="light green")
plot(u,e1,type="l",lty=1,ylab="",xlab="")
abline(0,0)
acf(e1,ylab="",xlab="",ylim=c(-0.5,1),lag=30,main="")
dev.off()

# Take different and fit a simple regression again
fit2=lm(y_diff~x_diff)            # Model 2
e2=fit2$resid

postscript(file="c:/res-teach/xiada/teaching05-07/figs/fig-2.4.eps",
horizontal=F,width=6,height=6)
matplot(u[-1],cbind(x_diff,y_diff),type="l",lty=c(1,2),col=c(1,2),
 ylab="",xlab="")
abline(0,0)
dev.off()

postscript(file="c:/res-teach/xiada/teaching05-07/figs/fig-2.5.eps",
horizontal=F,width=6,height=6)
par(mfrow=c(2,2),mex=0.4,bg="light pink")
ts.plot(e2,type="l",lty=1,ylab="",xlab="")
abline(0,0)
acf(e2, ylab="", xlab="",ylim=c(-0.5,),lag=30,main="")
# fit a model to the differenced data with an MA(1) error
fit3=arima(y_diff,xreg=x_diff, order=c(0,0,1))      # Model 3
e3=fit3$resid
```

```
ts.plot(e3,type="l",lty=1,ylab="",xlab="")
abline(0,0)
acf(e3, ylab="",xlab="",ylim=c(-0.5,1),lag=30,main="")
dev.off()
#################################################################

library(sandwich)  # HC and HAC are in the package "sandwich"
library(zoo)
z<-read.table("c:/res-teach/xiada/teaching05-07/data/ex2-1.txt",header=F)
x=z[,1]
y=z[,2]
x_diff=diff(x)
y_diff=diff(y)
# Fit a simple regression model and examine the residuals
fit1=lm(y_diff~x_diff)
print(summary(fit1))
e1=fit1$resid
# Heteroskedasticity-Consistent Covariance Matrix Estimation
#hc0=vcovHC(fit1,type="const")
#print(sqrt(diag(hc0)))
# type=c("const","HC","HC0","HC1","HC2","HC3","HC4")
# HC0 is the White estimator
hc1=vcovHC(fit1,type="HC0")
print(sqrt(diag(hc1)))
#Heteroskedasticity and autocorrelation consistent (HAC) estimation
#of the covariance matrix of the coefficient estimates in a
#(generalized) linear regression model.
hac1=vcovHAC(fit1,sandwich=T)
print(sqrt(diag(hac1)))
```

References

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817-858.

Cribari-Neto, F. and W. B. da Silva (2011). A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. *AStA Advances in Statistical*

Analysis, **95**, 129-146.

Eicker, F. (1967). Limit theorems for regression with unequal and dependent errors. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (L. LeCam and J. Neyman, eds.), University of California Press, Berkeley.

Newey, W.K. and K.D. West (1987). A simple, positive-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703-708.

Newey, W.K. and K.D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, **61**, 631-653.

White, H. (1980). A Heteroskedasticity consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, **48**, 817-838.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, Volume **11**, Issue 10.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, **16**, 1-16.