Lecture 8. Times Series Regression Models

1. Linear Regression

In many applications, the relationship between two time series, $\{x_t\}$ and $\{y_t\}$, is of interest. Consider a linear time series regression model in the form

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{t,2} + \cdots + \beta_K x_{t,K} + \varepsilon_t \\ &= x_t^{'}\beta + \varepsilon_t \end{aligned}$$

for $t = 1, \ldots, T$, where $x_t = \begin{bmatrix} 1 & x_{t,2} & \cdots & x_{t,K} \end{bmatrix}^{'}$ is a $K \times 1$ vector of explanatory variables, and $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_K \end{bmatrix}^{'}$ is a $K \times 1$ vector of coefficients, and $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

- Alternatively, one writes

$$y = X\beta + \varepsilon,$$

where $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_T \end{bmatrix}^{'}$ and $\varepsilon = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_T \end{bmatrix}^{'}$ are $T \times 1$ vectors and

$K \times 1$

$$X = \begin{bmatrix} x_1^{'} \\ x_2^{'} \\ \vdots \\ x_T^{'} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,2} & x_{1,3} & \cdots & x_{1,K} \\ 1 & x_{2,2} & x_{2,3} & \cdots & x_{2,K} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T,2} & x_{T,3} & \cdots & x_{T,K} \end{bmatrix} \qquad \underbrace{y}_{T\times 1} = \underbrace{X}_{T\times K}\overbrace{\beta} + \underbrace{\varepsilon}_{T\times 1}$$

$T \times K$

is a $T \times K$ matrix.

1.1. *Least Squares Estimation*

Ordinary least squares (*OLS*) estimation is based on minimizing

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - x_t^{'}\beta)^2$$

$X : T \times K$

$X' : K \times T$

with respect to β.

$X'X = K \times T \times T \times K$

- As a solution to the minimization problem, the OLS estimate b of β is given by $= K \times K$

$$b = (X^{'}X)^{-1}X^{'}y.$$

$K \times 1.$

$(X'X)^{-1}X' : K \times T$

For a given value of b, one estimates ε_t with the OLS residual $e_t = y_t - x_t^{'}b$. $b : K \times 1.$

Theorem 1.1. *Assume that (a) $E[x_t\varepsilon_t] = 0$, (b) $X^{'}X$ is of full rank K, and (c) $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$. Then, the followings are satisfied:*

↳ all column independent ⇒ invertible

1. *The OLS estimate b is unbiased: i.e., $E[b] = \beta$*
2. $b \sim N(\beta, \sigma_\varepsilon^2 (X^{'}X)^{-1})$

$Var[b] = \sigma_\varepsilon^2 (X'X)^{-1}$

$b = (X^TX)^{-1}X^Ty$

$b \sim N(\beta, \sigma_\varepsilon^2 (X^TX)^{-1})$

$S^2 = \frac{e'e}{T-K}$

3. *The unbiased estimate of σ^2 is given by*

$$s^2 = \frac{e'e}{T-K},$$

where $e = \begin{bmatrix} e_1 & e_2 & \cdots & e_T \end{bmatrix}'$.

$R^2 = 1 - \frac{e'e}{(y-\bar{y}i_T)'(y-\bar{y}i_T)}$

1.2. *Goodness of Fit*

Goodness of fit is summarized by the R^2 of the regression:

$R^2 = \frac{SSR}{TSS} = 1 - \frac{SSE}{TSS}$

$$R^2 = 1 - \frac{e'e}{(y-\bar{y}i_T)'(y-\bar{y}i_T)},$$

where $\bar{y} = (1/T)\sum_{t=1}^{T} y_t$ and $i_T = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}'$.

- R^2 measures the percentage of the variability of y_t that is explained by the regressors, x_t.

Remark 1.2. R^2 never decreases as more variables are added to the regression, even if the extra variables are irrelevant. To remedy this undesirable feature, the *adjusted* R^2 can be used instead.

1.3. *Hypothesis Testing*

Theorem 1.3. *Consider testing the hypothesis $H_0 : \beta_k = 0$ versus $H_1 : \beta_k \neq 0$. Under H_0, it shows*

$t_K = \frac{b_K}{\sqrt{S^2(X'X)^{-1}_{KK}}} \sim t(T-K)$

$$t_k = \frac{b_k}{\sqrt{s^2(X'X)^{-1}_{kk}}} \sim t_{(T-K)},$$

$H_0 : \beta_1 = \cdots \beta_K = 0.$
$H_1 : \beta \neq 0$

where $(X'X)^{-1}_{kk}$ denotes the kth diagonal element of $(X'X)^{-1}$.

rejection region

- If $|t_k| > t_{0.025}$, where $t_{0.025}$ is the 2.5% critical value from the t distribution with $T-K$ degrees of freedom, then H_0 is rejected at the 5% significance level. Equivalently, H_0 is rejected at the 5% significance level if the p-value is less than 0.05.

p-value

1.4. *Residual Diagnostics* 독립성, 정규성.

Residual diagnostic statistics evaluate the validity of the underlying assumptions of the model and serve as warning flags for possible misspecification.

residual
independent

- The *Durbin-Watson statistic* is given by

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}.$$

독립성 검정

$DW \approx 2$ indicates no serial correlation. If the 1st-order autocorrelation is positive, then $0 \leq DW < 2$. If the 1st-order autocorrelation is negative, then $2 < DW \leq 4$.

- The *Jarque-Bera test statistic* is given by

정규성 검정.

$$JB = \frac{T}{6}\left(\hat{S}^2(e) + \frac{(\hat{K}(e)-3)^2}{4}\right),$$

$\sim N$ or not

where $\hat{S}(e)$ and $\hat{K}(e)$ denote the sample skewness and kurtosis of e_t, respectively. Under the null hypothesis that e_t is normally distributed, it shows

$$JB \sim \chi^2_{(2)}.$$

1.5. Example: Capital Asset Pricing Model

The *Capital Asset Pricing Model* (CAPM) states that a linear relationship exists between the risk premium of any individual stocks, $E[r_i] - r_f$, and β_i in equilibrium:

$$E[r_i] - r_f = \beta_i (E[r_M] - r_f).$$

The CAPM defines the relationship between systematic risk and return, and quantifies this tradeoff.

- In the right-hand side of the CAPM, (a) β_i measures the amount of systematic risk of stock i relative to that of the market portfolio and (b) the risk premium of the market portfolio, $E[r_M] - r_f$, measures the reward for bearing the additional systematic risk if you hold the market portfolio. Thus, $\beta_i (E[r_M] - r_f)$ is the risk premium of holding the stock i relative to holding the riskless asset.

The CAPM can be estimated by running a time series regression of the form

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i (r_{M,t} - r_{f,t}) + \varepsilon_{i,t}$$

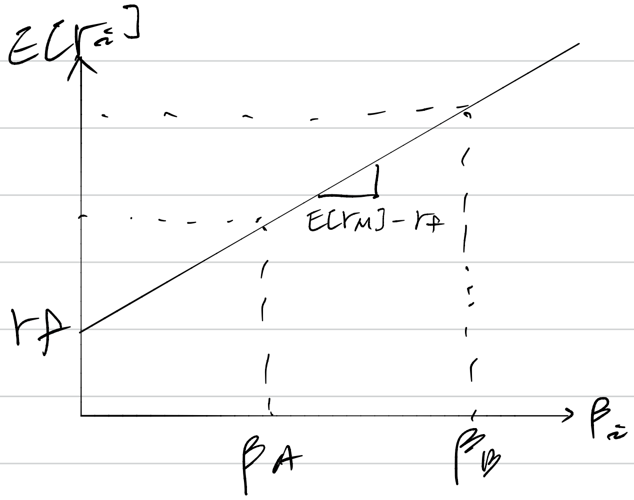
for $t = 1, \ldots, T$, where $r_{i,t}$ is the return on asset i, $r_{M,t}$ is the return on a market index portfolio, $r_{f,t}$ is the return on a risk-free asset, and $\varepsilon_{i,t} \sim i.i.d. N(0, \sigma^2_\varepsilon)$.

- In practice, the market index portfolio is some well diversified portfolios like the S&P 500 index. $r_{f,t}$ is taken as the T-bill rate to match the investment horizon associated with $r_{i,t}$.

- Suppose that the CAPM is "true." Then, the intercept α_i represents the actual performance of the stock relative to its predicted performance by the CAPM. The stock *outperforms* relative to its CAPM prediction if $\alpha_i > 0$, while the stock *underperforms* relative to its CAPM prediction if $\alpha_i < 0$.

Example 1.4. Consider the estimation of the CAPM regression for Microsoft using monthly return over the periods 4/1986 to 6/2012.

```
> mydat <- read.csv("data1.csv", header = T)
> head(mydat)
      Date     MSFT MKT_RF   RF  SMB   HML
1 4/1/1986 28.57143  -1.31 0.52 2.82 -2.91
...
6 9/2/1986  0.00000  -8.35 0.45 2.30  3.22
> r_rf <- ts(round(mydat$MSFT - mydat$RF, digits = 2), start = c(1986, 4),
     freq = 12)
```



observation

y   x

$r_i - r_f = \alpha_i + \beta_i (r_M - r_f) + \varepsilon_i$ risk = sys + unsys

$E[r_i] - r_f = \beta_i (E[r_M] - r_f)$ β ceo death.

$H_0: \alpha_i = 0$ $H_1: \alpha_i \neq 0$

If $\alpha_i > 0$, real observation

out perform than CAPM prediction

```
> mkt_rf <- ts(mydat$MKT_RF, start = c(1986, 4), freq = 12)
> fit <- lm(r_rf ~ mkt_rf)
> summary(fit)
              Estimate Std.Error  t value Pr(>|t|)
(Intercept)    1.5144    0.5041    3.004  0.00288 **
mkt_rf         1.2351    0.1075   11.493  < 2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

α : significant.
$\Rightarrow$ 1.5144
MSFT stock has outperformed relative to the CAPM prediction

- The estimated value for β for Microsoft is 1.2351, so Microsoft is judged to be riskier than the market index. The p-value for $H_0 : \alpha = 0$ is less than 0.05, meaning that α is significantly different from zero at the 5% level. It concludes that the Microsoft stock has outperformed relative to the CAPM prediction.

```
> e <- resid(fit)
> library(lmtest)
> dwtest(fit, alternative = "two.sided")
Durbin-Watson test
data: fit
DW = 2.0682, p-value = 0.5475
alternative hypothesis: true autocorrelation is not 0
```

H_0 : residual independent

- The p-value is no less than 0.05, concluding that the null of no autocorrelation is not rejected.

```
> library(tseries)
> jarque.bera.test(e)
Jarque Bera Test
data: e
X-squared = 56.6489, df = 2, p-value = 4.998e-13
```

H_0 : residual normality.

- The p-value is close to zero, so the null that residuals are normally distributed is rejected at the 5% level.

$e \nsim N$

2. Dynamic Regression Models : lead lag relation

The time series regression model may contain "lagged" variables as regressors to capture *dynamic effects*. The general dynamic time series regression model contains lagged values of y_t and lagged values of x_t:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \beta_0 x_t + \cdots + \beta_m x_{t-m} + \varepsilon_t,$$

old lead
new lag

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

ADL

- This model is called an *autoregressive distributed lag* (ADL) model and generalizes an AR(p) model for y_t by including m exogenous stationary regressors (i.e., $x_{t-1}, \ldots, x_{t-m}$).

외생

2.1. *Simple ADL Model*

Consider an ADL model with an exogenous variable x_t of the form

$= \sum_{i=0} (\phi_i y_{t-i} + \beta_i x_{t-i}) + \varepsilon_t$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \underbrace{\beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t}_{w_t}, \tag{2.1}$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ and $|\phi_1| < 1$. Let $w_t = \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$ and write (2.1) as

$$(1 - \phi_1 L) y_t = \phi_0 + w_t.$$

Then, one obtains

$$\begin{aligned}
y_t &= \frac{\phi_0}{1-\phi_1} + \frac{w_t}{1-\phi_1 L} \\
&= \frac{\phi_0}{1-\phi_1} + w_t + \phi_1 w_{t-1} + \phi_1^2 w_{t-2} + \cdots \\
&= \frac{\phi_0}{1-\phi_1} + (\beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t) + \phi_1(\beta_0 x_{t-1} + \beta_1 x_{t-2} + \varepsilon_{t-1}) \\
&\quad + \phi_1^2(\beta_0 x_{t-2} + \beta_1 x_{t-3} + \varepsilon_{t-2}) + \cdots \\
&= \frac{\phi_0}{1-\phi_1} + \beta_0 x_t + (\beta_1 + \phi_1 \beta_0) x_{t-1} + \phi_1(\beta_1 + \phi_1 \beta_0) x_{t-2} + \cdots \\
&\quad + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \cdots .
\end{aligned}$$

- The *immediate* impact multiplier (i.e., the impact of a change in x_t on y_t) is

$$\frac{\partial y_t}{\partial x_t} = \beta_0,$$

  the first lag multiplier (i.e, the impact of x_{t-1} on y_t) is

$$\frac{\partial y_{t+1}}{\partial x_t} = \frac{\partial y_t}{\partial x_{t-1}} = \beta_1 + \phi_1 \beta_0,$$

  and so on. In general, the kth lag multiplier is

$$\frac{\partial y_{t+k}}{\partial x_t} \quad \frac{\partial y_t}{\partial x_{t-k}} = \phi_1^{k-1}(\beta_1 + \phi_1 \beta_0).$$

If $\phi_1 y_{t-1} = 0$

$\frac{\partial y_t}{\partial x_{t-1}} = \beta_1$

$\frac{\partial y_t}{\partial x_{t-2}} = 0$

The *long-run effect* is defined by the cumulative sum of all the lag impact multipliers: i.e.,

$$\begin{aligned}
\text{long-run effect} &= \frac{\partial y_t}{\partial x_t} + \frac{\partial y_t}{\partial x_{t-1}} + \frac{\partial y_t}{\partial x_{t-2}} + \frac{\partial y_t}{\partial x_{t-3}} \cdots \\
&= \beta_0 + (\beta_1 + \phi_1 \beta_0) + \phi_1(\beta_1 + \phi_1 \beta_0) + \phi_1^2(\beta_1 + \phi_1 \beta_0) \cdots \\
&= (\beta_0 + \beta_1) + \phi_1(\beta_0 + \beta_1) + \phi_1^2(\beta_0 + \beta_1) + \cdots \\
&= \sum_{k=0}^{\infty} \phi_1^k (\beta_0 + \beta_1) \\
&= \frac{\beta_0 + \beta_1}{1 - \phi_1}.
\end{aligned}$$

Simple ADL model.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

where $|\phi_1| < 1$ $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, $w_t = \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$

$$(1 - \phi_1 L) y_t = \phi_0 + w_t$$

$$\begin{aligned} y_t &= \frac{\phi_0}{1-\phi_1} + \frac{w_t}{1-\phi_1 L} \\ &= \frac{\phi_0}{1-\phi_1} + w_t + \phi_1 w_{t-1} + \phi_1^2 w_{t-2} + \phi_1^3 w_{t-3} + \cdots \\ &= \frac{\phi_0}{1-\phi_1} + (\beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t) + \phi_1(\beta_0 x_{t-1} + \beta_1 x_{t-2} + \varepsilon_{t-1}) \\ &\quad + \phi_1^2(\beta_0 x_{t-2} + \beta_1 x_{t-3} + \varepsilon_{t-2}) + \cdots \\ &= \frac{\phi_0}{1-\phi_1} + \beta_0 x_t + (\beta_1 + \phi_1\beta_0) x_{t-1} + (\phi_1\beta_1 + \phi_1^2\beta_0) x_{t-2} + \\ &\quad + (\phi_1^2\beta_1 + \phi_1^3\beta_0) x_{t-3} + \cdots \\ &\quad + \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \cdots \\ &= \frac{\phi_0}{1-\phi_1} + \beta_0 x_t + (\beta_1 + \phi_1\beta_0) x_{t-1} + \phi_1(\beta_1 + \phi_1\beta_0) x_{t-2} \\ &\quad + \phi_1^2(\beta_1 + \phi_1\beta_0) x_{t-3} + \cdots \quad + \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} \end{aligned}$$

immediate impact multiplier

$$\frac{\partial y_t}{\partial x_t} = \beta_0$$

first lag (impact) multiplier

$$\frac{\partial y_t}{\partial x_{t-1}} = \beta_1 + \phi_1\beta_0$$

k-th lag (impact) multiplier

$$\frac{\partial y_t}{\partial x_{t-k}} = \phi_1^{k-1}(\beta_1 + \phi_1\beta_0)$$

long run effect

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{\partial y_t}{\partial x_{t-i}} &= \beta_0 + (\beta_1 + \phi_1\beta_0) + \phi_1(\beta_1 + \phi_1\beta_0) + \cdots \\ &= \beta_0 + (\beta_1 + \phi_1\beta_0) \cdot \frac{1}{1-\phi_1} \\ &= \frac{1}{1-\phi_1}(\beta_0 - \cancel{\beta_0\phi_1} + \beta_1 + \cancel{\phi_1\beta_0}) \\ &= \frac{1}{1-\phi_1}(\beta_0 + \beta_1) \end{aligned}$$

- The parameter ϕ_1 determines the *speed of adjustment*. If $\phi_1 = 0$, the long-run impact is reached in one time period, since

$$\begin{aligned} \text{long-run effect} &= \beta_0 + \beta_1 \\ &= \frac{\partial y_t}{\partial x_t} + \frac{\partial y_t}{\partial x_{t-1}}. \end{aligned}$$

If $\phi_1 \approx 1$, the long-run impact takes many periods, since $\partial y_t / \partial x_{t-k}$ is nonzero for many k.

2.2. *Example: Market Model*

$|\phi_1| \approx 1 \Rightarrow$ Speed of Conv. Slow

Consider an $N \times 1$ vector of asset return, denoted by r_t, follows the joint normal distribution: i.e.,

$$r_t = \begin{bmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{bmatrix} \sim MVN(\mu, \Sigma),$$

covariance matrix.

where $r_{i,t}$ is the return on asset i, μ is an $N \times 1$ mean vector, and Σ is an $N \times N$ variance matrix.

- For any security i, it shows

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \varepsilon_{i,t}, \tag{2.2}$$

where $r_{M,t}$ is the return on a market index portfolio and $\varepsilon_{i,t} \sim i.i.d.N(0, \sigma_\varepsilon^2)$.

Remark 2.1. A *market model* in the form of (2.2) assumes a linear relationship between the market return and the security return.

cf) CAPM assumes a linear relationship between risk premium of security and β_i (or market premium)

Example 2.2. Lo and MacKinlay (1990) show that small stocks tend to react with a week or more delay to common news. Consider a simple dynamic version of the market model

$$r_{i,t} = \phi_0 + \phi_1 r_{i,t-1} + \beta_0 r_{M,t} + \beta_1 r_{M,t-1} + \varepsilon_{i,t}$$

$r_{i,t} = \phi_0 + \phi_1 r_{i,t-1} + \beta_0 r_{M,t} + \beta_1 r_{M,t-1} + \varepsilon_{i,t}$

for weekly returns on Datalink Corporation over the periods 8/9/1999 to 10/8/2012. The NASDAQ composite index return is used as $r_{M,t}$.

```
> mydat <- read.csv("data2.csv", header = T)
> head(mydat)
       Date      DTLK     NASDAQ
1  8/9/1999  4.031209  3.5259442
...
6 9/13/1999 -9.564684 -0.6040747
> DTLK <- mydat[ , 2]
> nobs <- length(DTLK)
> DTLK.lag1 <- c(NA, DTLK[-nobs])
> NASDAQ <- mydat[ , 3]
> NASDAQ.lag1 <- c(NA, NASDAQ[-nobs])
```

```
> reg <- lm(DTLK ~ DTLK.lag1 + NASDAQ + NASDAQ.lag1)
> summary(fit)
             Estimate  Std.Error t value  Pr(>|t|)
(Intercept)  0.41356    0.36658   1.128  0.259648
DTLK.lag1   -0.08075    0.03793  -2.129  0.033614 *
NASDAQ       0.87460    0.09842   8.886   < 2e-16 ***
NASDAQ.lag1  0.39405    0.10371   3.799  0.000158 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The least squares estimates of ADL parameters are $\hat{\phi}_0 = 0.414$, $\hat{\phi}_1 = -0.081$, $b_0 = 0.875$, and $b_1 = 0.394$.

$\hat{r}_{i,t} = 0.414 - 0.081 r_{i,t-1} + 0.875 r_{M,t} + 0.394 r_{M,t-1}$

3. Linear Regression Models with Autocorrelated Errors

Consider a linear regression in the form

linear regression

$$y_t = x_t'\beta + u_t. \tag{3.1}$$

$y_t = x_t'\beta + \varepsilon_t$

where $x_t = [1\ x_{t,2} \cdots x_{t,k}]'$

If u_t is a white noise process, the OLS estimate b is consistent and efficient. If the error term u_t is serially correlated, however, b is consistent but "not" efficient.

not WN, $Cov(x_t, x_{t+k}) \neq 0$.

- For $u_t \sim WN(0, \sigma_u^2)$, $Var[b] = \sigma_u^2(X'X)^{-1}$. For autocorrelated u_t, $Var[b]$ is no longer equal to $\sigma_u^2(X'X)^{-1}$, and thus the statistical inference (based on $\hat{\sigma}_u^2(X'X)^{-1}$) may be misleading. There are two remedies: GLS approach and OLS approach.

3.1. *GLS Approach*

Suppose that u_t follows an ARMA(p, q) process with zero mean:

$$\phi(L)u_t = \theta(L)\varepsilon_t,$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Then, one writes (3.1) as

$$\phi(L)(y_t - x_t'\beta) = \theta(L)\varepsilon_t,$$

which can be regarded as an ARMA(p, q) model for a newly constructed time series $y_t - x_t'\beta$. Thus, one can estimate β using the ARMA model as usual. This approach is referred to as a special case of the *generalized least squares* (GLS) estimation.

- If errors are "correctly" specified, the GLS estimate is efficient. If the error structure is misspecified, however, the precision of the GLS estimate is often problematic; that is, the GLS approach is sensitive to the error specification.

3.2. OLS Approach

In the presence of autocorrelated errors, the OLS estimate is inefficient but "robust" to the error specification. This is a major advantage of the OLS approach in relative to the GLS approach. Based on this insight, an alternative estimation strategy in the presence of autocorrelated errors is to use OLS estimates with *heteroskedasticity and autocorrelation consistent* (HAC) variance matrix estimate.

- The popular HAC variance matrix estimate, due to Newey and West (1987), has the form

$$EstAsyVar_{HAC}[b] = (X'X)^{-1}\hat{S}_{HAC}(X'X)^{-1} \tag{3.2}$$

where

$$\hat{S}_{HAC} = \sum_{t=1}^{T} e_t^2 x_t x_t' + \sum_{l=1}^{q} w_l \sum_{t=l+1}^{T} (x_t e_t e_{t-l} x_{t-l}' + x_{t-l} e_{t-l} e_t x_t')$$

$y_t = \alpha + \beta x_t + u_t$

and w_l is the Bartlett weight function

u_t : OLS

$$w_l = 1 - \frac{l}{q+1}.$$

u_t SC ⇒ OLS b_k, se(b_k)

$\sqrt{\hat\sigma_u^2 (X'X)^{-1}_{HAC}}$

The HAC t-statistic is

$$t_k = \frac{b_k}{se_{HAC}(b_k)},$$

≠ GLS $\phi(L)(y_t - \alpha - \beta x_t) = \theta(L)\varepsilon_t$

where $se_{HAC}(b_k)$ is the square root of the diagonal elements of (3.2).

OLS w/ HAC $b = (X'X)^{-1}X'y$

3.3. Example: Long Run Regressions of Stock Returns on Dividend-Price Ratio

There has been much interest in whether long-term stock returns are predictable by valuation ratios like dividend-to-price and earnings-to-prices. The predictability is investigated by regressing future multiperiod stock returns on current values of valuation ratios.

- Let r_t denote the continuously compounded return on an asset in year t and let $d_t - p_t$ denote the log dividend price ratio (i.e., $\log(D_t/P_t) = \log(D_t) - \log(P_t)$). The predictive regression is given by

$$r_{t \to t+K} = \alpha + \beta(d_t - p_t) + \varepsilon_{t \to t+K} \tag{3.3}$$

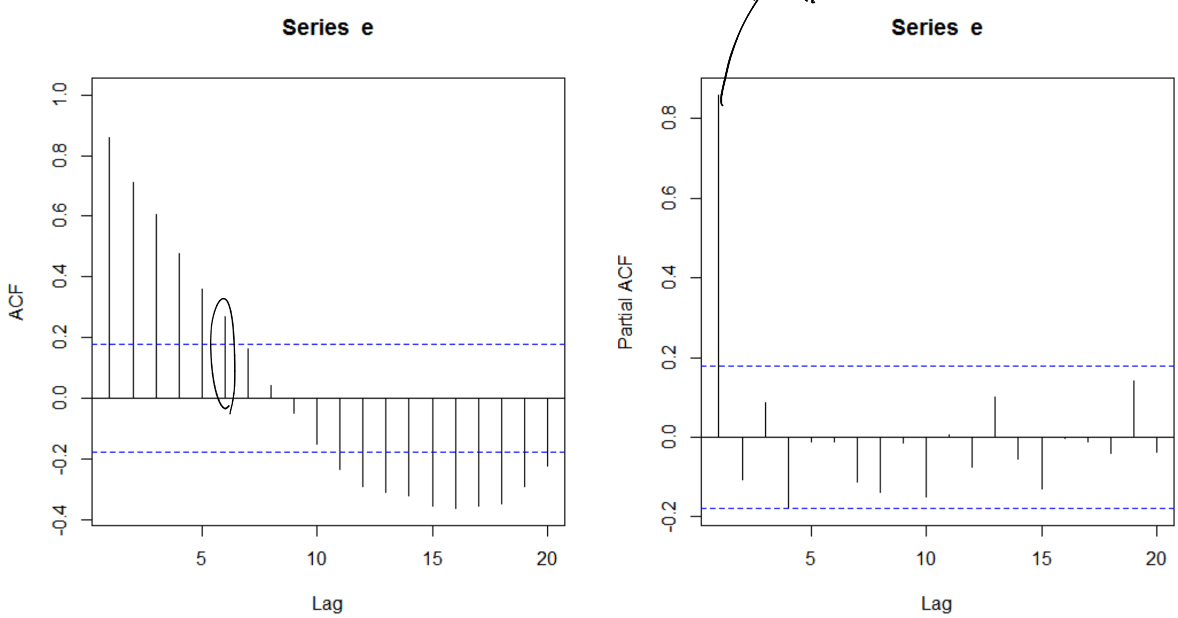
for $t = 1, \ldots, T$, where $r_{t \to t+K}$ is the continuously compounded future K-year return. The dividend-price ratio "predicts" future returns if $\beta \neq 0$.

The long-horizon regression with $K = 10$ years is estimated using the annual stock price and dividend data on the S&P 500 composite index from year 1871 to year 2000.

```
> mydat <- read.csv("data3.csv", header = T)
> head(mydat)
      ret10        dp
1 1.0443916 -2.903111
...
6 1.2838356 -2.479336
> ret10 <- mydat[ , 1]
> dp <- mydat[ , 2]
```

```
> reg1 <- lm(ret10 ~ dp)
> e <- resid(reg1)
> par(mfrow = c(1, 2))
> acf(e, lag = 20, xlim = c(1, 20))
> pacf(e, lag = 20)
```

AR(1).
= ARMA(1,0).



- The ACF shows that OLS residuals are serially correlated. Based on the PACF, one may choose an ARMA(1, 0) for u_t.

```
> datmat <- cbind(1, dp)
> (reg2 <- arima(ret10, order = c(1, 0, 0), xreg = datmat, include.mean = F,
     method = "ML"))
        ar1            dp
     0.8900  2.3887  0.5433
s.e. 0.0419  0.3245  0.0924
> tstat <- reg2$coef/sqrt(diag(reg2$var.coef))
> print(tstat, digits = 3)
  ar1        dp
21.24 7.36 5.88
```

$y_t = x_t'\beta + u_t$

$u_t \sim ARMA(1,0).$

$\phi(L)\, u_t = \theta(L)\, \varepsilon_t = \varepsilon_t$

$(1-\phi_1 L)(y_t - x_t'\beta) = \varepsilon_t$

$(1-\phi_1 L)(y_t - \phi_0 - \beta d_t) = \varepsilon_t$

- The fitted model is

$$(1-0.8900L)(y_t - 2.2887 - 0.5433x_t) = \varepsilon_t,$$

where y_t and x_t represent $r_{t\to t+K}$ and $d_t - p_t$, respectively, in (3.3). Since the t-statistic for β is 5.88, it concludes that future 10 year real total returns are highly predictable and positively related to the current dividend-price ratio.

```
> summary(reg1)
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.9457     0.4384   6.719 6.83e-10
dp           0.7475     0.1430   5.226 7.56e-07
> library(sandwich)
> nw.se <- sqrt(diag(NeweyWest(reg1)))
> (nw.t <- coef(reg1)/nw.se)
(Intercept)       dp
   3.096387 2.220827
```

- Given the serial correlation, the Newey-West HAC estimation provides a robust estimated standard error of the OLS b. The fitted model is

$$y_t = 2.9457 + 0.7475x_t + \hat{\varepsilon}_t$$

and the HAC consistent t-statistic for β is 2.2208. The predictability is confirmed again.