Regression Analysis and Forecasting : Chapter 3

I. Introduction

Simple linear regression : $y = \beta_0 + \beta_1 x + \varepsilon$

Multiple linear regression model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$.

Other examples: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$

$$y_t = \beta_0 + \beta_1 \sin \frac{2\pi}{d} t + \beta_2 \cos \frac{2\pi}{d} t + \varepsilon_t .$$

Cross-section data : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$, $i = 1, 2, \ldots, n$.

Time-series data : $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t$, $t = 1, 2, \ldots, T$.

II. Least-Squares Estimation

Let $\varepsilon_i \sim D(0, \sigma^2)$; $i = 1, 2, \ldots, n$, where $n > k$

Least squares f^n: $L = \sum_{i=1}^{n} \varepsilon_i^2$

$$= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2$$

$$= \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij}\right)^2$$

For a minima of $\{\hat{\beta}_j\}_{j=0,1,\ldots,k}$, FOC's are:

i) $$\left.\frac{\partial L}{\partial \beta_0}\right|_{\beta_1, \ldots, \beta_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij}\right) = 0$$

ii) $$\left.\frac{\partial L}{\partial \beta_j}\right|_{\beta_0, \beta_{p \neq j}} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij}\right) x_{ij} = 0 \; ; \quad j = 1, 2, \ldots, k.$$

Least Square Normal Equations:

Simplifying i) and ii), we have:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} \cdot x_{i1} + \dots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} x_{i1} = \sum_{i=1}^{n} y_i x_{i1}$$

$$\vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} x_{ik} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} \cdot x_{ik} + \dots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 = \sum_{i=1}^{n} y_i x_{ik}$$

Matrix form:

$$y = X\beta + \varepsilon, \text{ where}$$

y = $(n \times 1)$ vector of observations
X = $(n \times p)$ matrix of regressors (Model Matrix)
β = $(p \times 1)$ vector of regression coefficients
ε = $(n \times 1)$ vector of random errors

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

$$= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta$$

$$= y'y - 2\beta'X'y + \beta'X'X\beta.$$

WHY? $\beta'X'y_{(1\times1)} = y'X\beta_{(1\times1)}$

$$\left.\frac{\partial L}{\partial \beta}\right|_{\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

$$\Rightarrow \quad (X'X)\hat{\beta} = X'y.$$

$$\therefore \quad \boxed{\hat{\beta} = (X'X)^{-1} X'y}$$

Fitted Values : $\hat{y} = X\hat{\beta}$.

Residual : $e = y - \hat{y} = y - X\hat{\beta}$

$$\hat{\sigma}^2 = \frac{SS_E}{n-k-1} = \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{n-k-1} = \frac{SS_E}{n-p}$$

If the model is "correct", then $\hat{\beta}$ is an unbiased estimator of the model parameters β :

$E(\hat{\beta}) = \beta$. Also,

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

<u>Example</u> : <u>Trend Adjustment</u> :

1.) Fit a model with a linear time trend
2.) Subtract the fitted values from original observations
3.) Now, residuals are <u>trend free</u> .
4.) Forecast residuals
5.) Add residual forecast value to estimate of trend, so
6.) Now, we have a forecast of ~~origin~~ the variable.

$$y_t = \beta_0 + \beta_1 t + \varepsilon \quad , \quad t = 1, 2, \dots, T.$$

Least squares normal equations for this model:

$$T\hat{\beta}_0 + \hat{\beta}_1 \frac{T(T+1)}{2} = \sum_{t=1}^{T} y_t$$

$$\hat{\beta}_0 \frac{T(T+1)}{2} + \hat{\beta}_1 \frac{T(T+1)(2T+1)}{6} = \sum_{t=1}^{T} t \cdot y_t .$$

$$\therefore \quad \hat{\beta}_0 = \frac{2(2T+1)}{T(T-1)} \sum_{t=1}^{T} y_t - \frac{6}{T(T-1)} \sum_{t=1}^{T} t\, y_t , \text{ and}$$

$$\hat{\beta}_1 = \frac{12}{T(T^2-1)} \sum_{t=1}^{T} t\, y_t - \frac{6}{T(T-1)} \sum_{t=1}^{T} y_t .$$

Procedure of previous page implies:

- Let $\hat{\beta}_0(T)$ and $\hat{\beta}_1(T)$ denote the estimates of parameters computed at point in time T.
- Predicting the next observation $\Rightarrow$
  a) predict the point on the trend line in period $T+1$ } $\hat{\beta}_0(T) + \hat{\beta}_1(T) * (T+1)$
  b) add a forecast of the next residual } $\hat{e}_{T+1}(1)$ ← should be zero if structureless.

$$\therefore \quad \hat{y}_{T+1}(T) = \hat{\beta}_0(T) + \hat{\beta}_1(T) * (T+1).$$

III. Statistical Inference

- Hypothesis testing
- Confidence Interval Estimation
- Assume ε_i (or ε_t) $\sim NID(0, \sigma^2)$

1. Test for significance of regression:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$
$$H_1: \text{at least one } \beta_j \neq 0$$

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = y'y - n\bar{y}^2$$

$$SSR = \hat{\beta}' X'y - n\bar{y}^2$$
$$SSE = y'y - \hat{\beta}' X'y$$
$$F_0 = \frac{SSR/k}{SSE/n-p}$$

ANOVA Table

Source of Variation	Sum of Squares	D.F.	Mean Square	Test Stat, F_0
Regression	SSR	k	$\frac{SSR}{k}$	$F_0 = \frac{SSR/k}{SSE/n-p}$
Residual / Error	SSE	$n-p$	$\frac{SSE}{n-p}$	
Total	SST	$n-1$		

Also see R^2 and Adjusted-R^2.

2. <u>Tests for significance of individual regression coefficients</u>

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

$$t_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Reject H_0 if $|t_0| > t_{\frac{\alpha}{2}, n-p}$

To test the significance of a group of coefficients, we can "partition" the model, and do a <u>partial F-test</u>.

Let $y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$.

Then, $H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

$$SSR(\beta_1 | \beta_2) = SSR(\beta) - SSR(\beta_2)$$
$$= \hat{\beta}' X' y - \hat{\beta}_2' X_2' y$$

$$\therefore \quad F_0 = \frac{SSR(\beta_1|\beta_2)/r}{\hat{\sigma}^2} \sim F(r, n-p)$$

3. <u>Confidence Intervals on individual regression coefficients</u>

If $\varepsilon \sim NID(0, \sigma^2)$, then

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

Let C_{jj} = $(jj)^{th}$ element of $(X'X)^{-1}$.
Then, a $(100-\alpha)\%$ C.I. for β_j, $j=0,1,\dots,k$ is:

$$\hat{\beta}_j - t_{\frac{\alpha}{2},n-p} \cdot \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\frac{\alpha}{2},n-p}\sqrt{\hat{\sigma}^2 C_{jj}}$$

or:

$$\hat{\beta}_j - t_{\frac{\alpha}{2},n-p} \cdot se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{\frac{\alpha}{2},n-p} \cdot se(\hat{\beta}_j)$$

4. <u>Confidence Interval on the Mean Response</u>

Let $x_0 = \begin{bmatrix} 1 \\ x_{01} \\ \vdots \\ x_{0k} \end{bmatrix}$ be a particular combination of regressors.

Then, <u>mean response</u> at this point is:

$$E[y(x_0)] = \mu_{y|x_0} = x_0' \beta \text{ , and}$$

$$\hat{y}(x_0) = \hat{\mu}_{y|x_0} = x_0' \hat{\beta} \text{ , and}$$

$$Var(\hat{y}(x_0)) = \hat{\sigma}^2 x_0' (X'X)^{-1} x_0$$

Standard error of fitted response,

$$s.e(\hat{y}(x_0)) = \sqrt{Var(\hat{y}(x_0))} = \sqrt{\hat{\sigma}^2 x_0'(X'X)^{-1}x_0}.$$

Then, a $(100-\alpha)\%$ C.I. for mean response at x_0 is:

$$\hat{y}(x_0) - t_{\frac{\alpha}{2},n-p} \cdot s.e(\hat{y}(x_0)) \leq \mu_{y|x_0} \leq \hat{y}(x_0) + t_{\frac{\alpha}{2},n-p} \cdot se(\hat{y}(x_0)).$$

5. Prediction of new observations

Let x_0 be a particular set of values of regressors.

Point estimate of future observation, $y(x_0)$ is :

$$\hat{y}(x_0) = x_0' \hat{\beta}$$

Prediction Error, $e(x_0) = y(x_0) - \hat{y}(x_0)$, and

$$Var(e(x_0)) = Var(y(x_0) - \hat{y}(x_0))$$

assuming independence b/w y & $\hat{y}$

$$= Var(y(x_0)) + Var(\hat{y}(x_0))$$
$$= \sigma^2 + \sigma^2 x_0' (X'X)^{-1} x_0$$
$$= \sigma^2 (1 + x_0'(X'X)^{-1} x_0)$$

$\therefore$ s.e. of prediction error $= \sqrt{\hat{\sigma}^2 (1 + x_0'(X'X)^{-1} x_0)}$
$\sim t_{n-p}$.

$\therefore$ A $(100-\alpha)$ prediction interval for $y(x_0)$ is :

$$\hat{y}(x_0) - t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 (1 + x_0'(X'X)^{-1} x_0)} \leq y(x_0) \leq$$
$$\hat{y}(x_0) + t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 (1 + x_0'(X'X)^{-1} x_0)}$$

CI : Interval estimate on MEAN RESPONSE of 'y' dist at x_0.

PI : Interval estimate on a single future observation from the y-dist at x_0.

6. Model Adequacy Checking :

A. Residual plots : $e_i = y_i - \hat{y}_i \; ; \; i = 1, 2, \ldots, n.$

Assumption is $\varepsilon_i \sim NID(0, \sigma^2)$.

Normal prob. plot : Normality of errors

Residuals vs. Fitted : Constant Variance | Equality of variance.

Residuals vs. Regressors

Residuals vs. time order : Esp. for time-series data.

B. Standardized Residuals : $d_i = \dfrac{e_i}{\hat{\sigma}} \; ; \; \hat{\sigma}^2 = MSE$

If $|d_i| > 3$, examine the outliers

C. Studentized Residuals : For heteroskedastic variance,

$$\hat{y} = X\hat{\beta} = X\left[(X'X)^{-1}X'y\right]$$
$$= (X(X'X)^{-1}X') \cdot y$$
$$= \underbrace{H}_{\text{Hat Matrix}} y$$

$$\therefore \; e = y - \hat{y} = y - Hy = y(I - H)$$

$Cov(e) = \sigma^2 (I - H)$: Covariance matrix.

$V(e_i) = \sigma^2 (1 - h_{ii})$

$\therefore$ Studentized Residuals : $r_i = \dfrac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}} \; ; \; \hat{\sigma}^2 = MSE$

D. <u>Prediction error sum of squares</u> (PRESS) DIY

E. <u>R-Student</u> : Externally studentized residual DIY

F. <u>Cook's Distance</u> : $D_i = \frac{r_i^2}{p} \cdot \frac{h_{ii}}{1-h_{ii}}$

If $D_i > 1 \Rightarrow$ Observation has influence!

<u>Note</u>: Also See Variable Selection Methods and model selection criteria. We will use 'R' to study this.

7. <u>Generalized least squares, GLS</u>

Problem: Heteroskedasticity or non-constant variance.
Solution(s): Transform 'y' variable

Suppose $V = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$, where

σ_i^2 = variance of i'th observation, y_i, $i=1,2,\ldots,n$.

We can create a "Weighted Least Squares" (WLS) fn, where the weight is inversely proportional to the variance of $y_i \Rightarrow w_i = \frac{1}{\sigma_i^2}$

Then, $L = \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2$, and the least squares normal equations are :

$$\hat{\beta}_0 \sum_{i=1}^{n} w_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} w_i x_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i^2 = \sum_{i=1}^{n} w_i x_i y_i$$

Solve these to get $\hat{\beta}_0$ and $\hat{\beta}_1$.

<u>In general,</u> if $y = X\beta + \varepsilon$, $E(\varepsilon) = 0$; and $Var(\varepsilon) = \sigma^2 V$ (not $\sigma^2 I$)

Then, the least squares ~~normal equations are~~ function is :

$$L = (y - X\beta)' V^{-1} (y - X\beta).$$

~~A~~ Least squares normal equations are :

$$(X' V^{-1} X) \hat{\beta}_{GLS} = X' V^{-1} y.$$

GLS estimator of β is :

$$\boxed{\hat{\beta}_{GLS} = (X' V^{-1} X)^{-1} X' V^{-1} y}$$, and

$$Var(\hat{\beta}_{GLS}) = \sigma^2 (X' V^{-1} X)^{-1}$$

Note : $\hat{\beta}_{GLS}$ is BLUE of β.

__Question__: How do we find the "weights"?
estimate

Steps in finding a "Variance equation":

1. Fit an OLS: $y = X\beta + \varepsilon$; obtain e (as residuals).
2. Use Residual diagnostics to determine if $\sigma_i^2 = f(y)$ or $\sigma_i^2 = f(x)$
   Resid. vs. fitted plot / Resid. vs. Regressor plot.
3. Regress $\sqrt{e}$ on either y or x_i, whichever is appropriate.
   Obtain an equation for predicting the variance of each observation: $\hat{s}_i^2 = f(x)$ or $\hat{s}_i^2 = f(y)$.
4. Use fitted values from the estimated variance f^n to obtain weights: $w_i = \frac{1}{\hat{s}_i^2}$; $i = 1, 2, \ldots, n$.
5. Use these weights as the diagonal elements of the V^{-1} matrix in the GLS procedure.
6. __Iterate__ to reduce difference b/w $\hat{\beta}_{OLS}$ & $\hat{\beta}_{GLS}$.

Example: See R code for example.

We can also use "Discounted" least squares, where recent observations are weighted more heavily than older observations.

8. Regression Models for general time-series data

If errors are correlated or not independent:

i) $\hat{\beta}_{OLS}$ is still unbiased, but not BLUE, i.e., $\hat{\beta}_{OLS}$ does not have minimum variance. [SHOW ... exercise].

ii) If errors are positively autocorrelated, MSE is an (serious) underestimate of σ^2.
=> s.e. of $\hat{\beta}$ are underestimated.
=> C.I's and P.I's are shorter than they should be.
=> H_0 may be rejected more frequently than we should be, and hypothesis tests are no longer reliable.

Attention
Solutions to Autocorrelated errors

a) If autocorrelation due to omitted variable, then identifying and including the appropriate omitted variable should remove the autocorrelation.

b) If we "understand" the structure of the autocorrelation, then use GLS

c) Use models that specifically accounts for the autocorrelation

Consider a simple linear regression with "first-order" autoregressive errors, i.e:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad , \text{ where}$$

$$\varepsilon_t = \phi \varepsilon_{t-1} + a_t \, , \quad a_t \sim NID(0, \sigma_a^2),$$
and $|\phi| < 1$.

First, note that:

$$\begin{aligned}
\varepsilon_t &= \phi \varepsilon_{t-1} + a_t \\
&= \phi(\phi \varepsilon_{t-2} + a_{t-1}) + a_t \\
&= \phi(\phi(\phi\varepsilon_{t-3} + a_{t-2}) + a_{t-1}) + a_t \\
&\vdots
\end{aligned}$$

$$\Rightarrow \varepsilon_t = \phi^2 \varepsilon_{t-2} + \phi a_{t-1} + a_t$$

$$\begin{aligned}
&= \phi^3 \varepsilon_{t-3} + \phi^2 a_{t-2} + \phi a_{t-1} + a_t \\
&= \phi^3 \varepsilon_{t-3} + \phi^2 a_{t-2} + \phi^1 a_{t-1} + \phi^0 a_t \\
&\vdots
\end{aligned}$$

$$\Rightarrow \varepsilon_t = \sum_{j=0}^{t} \phi^j a_{t-j}$$

Also : [SHOW THESE : DIY]

ⓐ $E(\varepsilon_t) = 0$ ~~show~~

ⓑ $Var(\varepsilon_t) = \sigma^2 = \sigma_a^2 \left(1/_{1-\phi^2} \right)$, and

ⓒ $Cov(\varepsilon_t, \varepsilon_{t \pm j}) = \phi^j \sigma_a^2 \left(1/_{1-\phi^2} \right)$.

Lag One autocorrelation, ρ_1 is:

$$\rho_1 = \frac{\text{cov}(\varepsilon_t, \varepsilon_{t+1})}{\sqrt{\text{Var}(\varepsilon_t)} \cdot \sqrt{\text{Var}(\varepsilon_{t+1})}} = \phi \quad \text{(SHOW)}$$

Lag 'k' autocorrelation, ρ_k is:

ACF: $$\rho_k = \frac{\text{cov}(\varepsilon_t, \varepsilon_{t+k})}{\sqrt{\text{Var}(\varepsilon_t)} \sqrt{\text{Var}(\varepsilon_{t+k})}} = \phi^k$$

1. <u>Durbin - Watson Test for autocorrelation</u>

Positive Autocorrelation: $H_0 : \phi = 0$
$H_1 : \phi > 1$ ← Correct

$$d = \sum_{t=2}^{T} (e_t - e_{t-1})^2 \Big/ \sum_{t=2}^{T} e_t^2 \approx 2 \cdot (1 - r_1),$$

where r_1 = lag one autocorrelation b/w residuals.

Positive Autocorrelation

If $d < d_L$: reject $H_0 : \phi = 0$
If $d > d_U$: do not reject H_0
If $d_L < d < d_U$: Inconclusive

Negative autocorrelation

$H_1 : \phi < 1$

If $(4-d) < d_L$: reject H_0
If $(4-d) > d_L$: do not reject H_0
If $d_L < 4-d < d_U$: Inconclusive.

2. Estimating parameters in a time series regression model.

Omission of one or more important predictor variables can cause 'artificial' time dependence.
See Example 3.13

Cochrane–Orcutt Method: Simple linear regression with AR(1) errors

Let $$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad ; \quad \varepsilon_t = \phi\,\varepsilon_{t-1} + a_t$$
$$a_t \sim NID(0, \sigma_a^2); \ |\phi| < 1$$

Step 1: Transform y_t : $y_t' = y_t - \phi\, y_{t-1}$

$$\Rightarrow \quad y_t' = \underbrace{\beta_0 + \beta_1 x_t + \varepsilon_t}_{y_t} - \phi\,(\underbrace{\beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}}_{y_{t-1}})$$

$$\Rightarrow \quad y_t' = \underbrace{\beta_0(1-\phi)}_{\beta_0'} + \beta_1(\underbrace{x_t - \phi\, x_{t-1}}_{x_t'}) + (\underbrace{\varepsilon_t - \phi\,\varepsilon_{t-1}}_{a_t})$$

$$(*) \quad \Rightarrow \quad y_t' = \beta_0' + \beta_1 x_t' + a_t$$

Step 2: Get e_t : $y_t - \hat{y}_t$

Step 3: Estimate ϕ : $e_t = \hat{\phi}\, e_{t-1}$

Recall : $$\hat{\phi} = \sum_{t=2}^{T} e_t e_{t-1} \Big/ \sum_{t=1}^{T} e_t^2$$ (Recall : $\frac{X'y}{X'X}$.

Step 4: Calculate $$y_t' = y_t - \hat{\phi}\, y_{t-1}$$
$$x_t' = x_t - \hat{\phi}\, x_{t-1}$$

Step 5: Estimate ⊛ : $\hat{y}_t' = \hat{\beta}_0' + \hat{\beta}_1 x_t'$

Step 6: Apply Durbin - Watson Test on $a_t = y_t' - \hat{y}_t'$.
If autocorrelation persists, REPEAT process!

See Example 3.14.

Maximum Likelihood Approach

Let $y_t = \mu + a_t$; $a_t \sim N(0, \sigma^2)$; μ = unknown constant.

Probability density f^n of $\{y_t\}_{t=1,2,\dots,T}$:

$$f(y_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y_t-\mu}{\sigma}\right)^2\right]$$

$$\Rightarrow \quad f(y_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{a_t}{\sigma}\right)^2\right]$$

Joint probability density f^n of $\{y_1, y_2, \dots, y_T\}$:

$$l(y_t, \mu) = \prod_{t=1}^{T} f(y_t) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{a_t}{\sigma}\right)^2\right]$$

$$\therefore \quad l(y_t, \mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T \exp\left[\frac{-1}{2\sigma^2} \sum_{t=1}^{T} a_t^2\right]$$

This is the likelihood f^n ↑

Log-likelihood f^n:

$$\ln(y_t, \mu) = -\frac{T}{2}\ln(2\pi) - T\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} a_t^2 .$$

To maximize the log-likelihood, minimize $\sum_{t=1}^{T} a_t^2$, i.e;

$$\text{Minimize } \sum_1^T a_t^2 = \sum_1^T (y_t - \mu)^2$$

This is the same as the LS estimators!!

New Consider $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$; $\varepsilon_t = \phi\varepsilon_{t-1} + a_t$;
$|\phi| < 1$; $a_t \sim N(0, \sigma_a^2)$

Then,

$$y_t - \phi y_{t-1} = (1-\phi)\beta_0 + \beta_1(x_t - \phi x_{t-1}) + a_t$$

(as in Cochrane-Orcutt)

$$\Rightarrow y_t = \phi y_{t-1} + (1-\phi)\beta_0 + \beta_1(x_t - \phi x_{t-1}) + a_t .$$

The joint pdf of a_t's is:

$$f(a_2, a_3, \ldots, a_T) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left[-\frac{1}{2}\left(\frac{a_t}{\sigma_a}\right)^2\right]$$

$$= \left[\frac{1}{\sqrt{2\pi\sigma_a^2}}\right]^{T-1} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=2}^{T} a_t^2\right]$$

Likelihood f^n: $l(y_t, \phi, \beta_0, \beta_1) =$

$$\left[\frac{1}{\sqrt{2\pi\sigma_a^2}}\right]^{T-1} * \exp\left[\frac{-1}{2\sigma_a^2}\sum_{t=2}^{T}\left\{y_t - (\phi y_{t-1} + (1-\phi)\beta_0 + \beta_1(x_t - \phi x_{t-1}))\right\}^2\right]$$

Log-likelihood : $\ln(y_t, \phi, \beta_0, \beta_1) =$

$$-\frac{T-1}{2}\ln(2\pi) - (T-1)\ln(\sigma_a)$$
$$-\frac{1}{2\sigma_a^2}\sum_{t=2}^{T}\left[y_t - \left(\phi y_{t-1} + (1-\phi)\beta_0 + \beta_1(x_t - \phi x_{t-1})\right)\right]^2.$$

To maximize log-likelihood, minimize last term of RHS, which is just the SSE !

<u>One-period ahead forecast</u> :

A $(100-\alpha)\%$ PI for lead-one forecast is :

$$\hat{y}_{T+1}(T) \pm z_{\frac{\alpha}{2}} \cdot \hat{\sigma}_a \quad , \quad \text{where}$$

$$\hat{y}_{T+1}(T) = \hat{\phi} y_T + (1-\hat{\phi})\hat{\beta}_0 + \hat{\beta}_1(x_{T+1} - \hat{\phi} x_T)$$

<u>Two-periods ahead forecast</u> :

$$\hat{y}_{T+2} = \hat{\phi}\hat{y}_{T+1} + (1-\hat{\phi})\hat{\beta}_0 + \hat{\beta}_1(x_{T+2} - \hat{\phi} x_{T+1})$$
$$= \hat{\phi}\left[\hat{\phi} y_T + (1-\hat{\phi})\hat{\beta}_0 + \hat{\beta}_1(x_{T+1} - \hat{\phi} x_T)\right]$$
$$+ (1-\hat{\phi})\hat{\beta}_0 + \hat{\beta}_1(x_{T+2} - \hat{\phi} x_{T+1}).$$

This is assuming, y_T and x_T are known ;
x_{T+1} and x_{T+2} are known ;
but a_{T+1} and a_{T+2} are not known (realized) yet.

Two-step-ahead forecast error:

$$y_{T+2} - \hat{y}_{T+2}(T) = a_{T+2} + \hat{\phi}\, a_{T+1}.$$

$$\begin{aligned} Var\left(a_{T+2} + \hat{\phi}\, a_{T+1}\right) &= \hat{\sigma}_a^2 + \hat{\phi}^2 \hat{\sigma}_a^2 \\ &= \left(1+\hat{\phi}^2\right)\hat{\sigma}_a^2 \end{aligned}$$

A $(100-\alpha)\%$ P.I. for two-step-ahead forecast is:

$$\hat{y}_{T+2}(T) \pm z_{\frac{\alpha}{2}} \cdot \left(1+\hat{\phi}^2\right)\hat{\sigma}_a.$$

<u>τ-periods-ahead forecasts</u>:

$$\hat{y}_{T+\tau}(T) = \hat{\phi}\, \hat{y}_{T+\tau-1}(T) + (1-\hat{\phi})\hat{\beta}_0 + \hat{\beta}_1\left(x_{T+\tau} - \hat{\phi}\, x_{T+\tau-1}\right)$$

τ-step-ahead forecast error:

$$y_{T+\tau} - \hat{y}_{T+\tau}(T) = a_{T+\tau} + \phi\, a_{T+\tau-1} + \dots + \phi^{\tau-1} a_{T+1}.$$

$$\begin{aligned} Var\left(y_{T+\tau} - \hat{y}_{T+\tau}(T)\right) &= \left(1+\phi^2+\phi^3+\dots+\phi^{2(\tau-1)}\right)\sigma_a^2 \\ &= \sigma_a^2\left(1-\phi^{2\tau}\right)/\left(1+\phi^{2}\right) \end{aligned}$$

$\therefore$ A $(100-\alpha)\%$ P.I. for the lead-τ forecast is:

$$\hat{y}_{T+\tau}(T) \pm z_{\frac{\alpha}{2}} \cdot \left(\frac{1-\hat{\phi}^{2\tau}}{1-\hat{\phi}^2}\right)^{1/2} \cdot \hat{\sigma}_a$$

<u>What if $X_{T+\tau}$ is not known?</u>

Use an unbiased forecast of x_{T+1} : $\hat{x}_{T+1}$, and use for forecast of $y_{T+1}(T)$, and to calculate variance of the forecast error: $\dfrac{1-\phi^{2\tau}}{1+\phi^2}\sigma_a^2 + \beta_1^2\, \sigma_x^2(\tau)$! See equations <u>3.114 – 3.118</u>.