Final

E4.27
CS5.2

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2010

MSc and EEE/ISE PART IV: MEng and ACGI

IDENTIFICATION AND ADAPTIVE CONTROL

Tuesday, 4 May 10:00 am

Time allowed: 3:00 hours

There are FIVE questions on this paper.

Answer FOUR questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : R.B. Vinter
Second Marker(s) : S. Evangelou

Information for candidates:

The Multivarite Normal Density:

The probability density $N(\mathbf{m},Q)$ of an n-vector, normal random variable with mean $\mathbf{m}$ and covariance matrix Q ($Q>0$) is

$$N(m,Q)(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det Q)^{\frac{1}{2}}}\exp -\frac{1}{2}\left\{(\mathbf{x}-\mathbf{m})^T Q^{-1}(\mathbf{x}-\mathbf{m})\right\} .$$

In the case that $n=1$, m is a scalar and $Q=\sigma^2$ ($\sigma^2>0$),

$$N(m,\sigma^2)(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

and, if X is a scalar random variable with probability density $N(m,\sigma^2)$,

$$\text{Prob}\{m-2\sigma \;\leq\; X \;\leq m+2\sigma\} \approx 0.95 .$$

The Cramer-Rao Lower Bound:

Take a family of probability densities $\{p(\mathbf{y};\theta)\}$ parameterised by the k-vector θ. Let $\hat{\theta}(\mathbf{y})$ be any unbiased estimate of θ given $\mathbf{y}$. Then the covariance of $\hat{\theta}(\mathbf{y})$ satisfies

$$cov\{\hat{\theta}(\mathbf{y})\} \geq M^{-1}(\theta)$$

where $M(\theta)$ is the $k\times k$ Fisher Information Matrix, with components $\{m_{ij}\}$ defined by:

$$m_{ij} = -E_\theta\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}log_e f(\mathbf{y},\theta)\} .$$

1. a) Consider the stationary processes $\{\mathbf{x}_t\}$ and $\{\mathbf{y}_t\}$ generated by the stochastic state space model

$$
\begin{aligned}
\mathbf{x}_{t+1} &= A\mathbf{x}_t + \mathbf{b}v_t \\
\mathbf{y}_t &= C\mathbf{x}_t \,.
\end{aligned}
$$

Here A is an $n \times n$ matrix with eigenvalues located strictly in the unit disc, C is an $r \times n$ matrix, $\mathbf{b}$ is an n-vector, and $\{v_t\}$ is a scalar white noise process with variance σ^2.

Derive equations for the covariance functions $R_x(k)$ and $R_y(k)$ of $\{\mathbf{x}_t\}$ and $\{\mathbf{y}_t\}$ respectively. [12]

b) Two correlated stationary processes $\{y_t^1\}$ and $\{y_t^2\}$ are described by the equations:

$$
\begin{aligned}
y_{t+1}^1 &= y_t^2 + v_t \\
y_{t+1}^2 &= -\alpha y_t^1 + v_t \,,
\end{aligned}
$$

in which v_t is a white noise process with variance σ^2. α is an unknown positive parameter. Suppose that

$$
\frac{R_{y^2(0)}}{R_{y^1(0)}} = 0.75 \,.
$$

Determine the value of α. [8]

Hint: use the results of part (a).

2. The scalar data sequence $\{y_1,\dots,y_N\}$ is associated with the model

$$y_i = x_i\theta + e_i \quad \text{for } i = 1,\dots,N,$$

in which $\{x_1,\dots,x_N\}$ is a given sequence of numbers , and $\{e_i\}$ is a sequence of independent scalar random variables such that

$$e_i \sim N(d,\sigma^2) \quad \text{for } i = 1,\dots,N.$$

Here, θ, d and $\sigma^2 (> 0)$ are parameters.

a) Assume that $d = 0$. Show that

$$\hat{\sigma}^2 = (N-1)^{-1}\sum_{i=1}^{N}|y_i - x_i\hat{\theta}|^2$$

is an unbiased estimate of σ^2, where $\hat{\theta}$ is the least squares estimate of θ. [8]

b) Now assume that σ^2 is known. Show that the least squares estimates of θ and d are

$$\hat{\hat{\theta}} = \frac{\hat{r}_{xy}}{\hat{r}_x} \quad \text{and} \quad \hat{\hat{d}} = \frac{\hat{r}_x m_y - \hat{r}_{xy} m_x}{\hat{r}_x},$$

where m_x, m_y, $\hat{r}_x$ and $\hat{r}_{xy}$ are the sample means and covariances:

$$m_x = (1/N)\sum_{i=1}^{N} x_i,\ m_y = (1/N)\sum_{i=1}^{N} y_i$$
$$\hat{r}_x = (1/N)\sum_{i=1}^{N}(x_i - m_x)^2,\ \hat{r}_{xy} = (1/N)\sum_{i=1}^{N}(x_i - m_x)(y_i - m_y).$$

[8]

Derive formulae for the error variances $E[(\theta - \hat{\hat{\theta}})^2]$ and $E[(d - \hat{\hat{d}})^2]$. [4]

3. a) An N-vector random variable $\mathbf{y}$ is modelled as

$$\mathbf{y} = \mathbf{x}\theta + \mathbf{e}\,, \tag{3.1}$$

where $\mathbf{x}$ is a given N-vector, θ is an unknown scalar parameter and $\mathbf{e}$ is a zero-mean Gaussian random variable with given covariance

$$cov\{\mathbf{e}\} = Q\,.$$

Let $\hat{\theta}$ be the weighted linear least squares estimate of θ given $\mathbf{y}$:

$$\hat{\theta} = (x^T Q^{-1} x)^{-1} x^T Q^{-1} \mathbf{y}\,.$$

Show that $\hat{\theta}$ is an unbiased estimate of θ given $\mathbf{y}$, whose error covariance achieves the Cramer Rao lower bound. [8]

Now consider the following random variable

$$z = a_0\theta + v\,,$$

in which a_0 is a given number, v is a zero mean, scalar Gaussian random variable, independent of $\mathbf{e}$, with given variance σ_0^2, and θ is the same unknown constant as above. We can use the earlier estimate $\hat{\theta}$ of θ to obtain the following estimate $\hat{z}$ of z given $\mathbf{y}$:

$$\hat{z} = a_0\hat{\theta}\,.$$

Determine a 95% confidence interval for z given $\mathbf{y}$, based on the estimate $\hat{z}$. (The endpoints of the interval will depend on Q, a_0 and σ_0^2.) [4]

b) Let $e_1, \ldots, e_N$ be a finite sequence of of scalar random variables, $N >> 1$. Describe a 'large sample' test for assessing whiteness of the sequence. Your description should include a summary of the properties of a white noise sequence on which the test is based. [8]

4. Consider the stationary process $\{y_t\}$ generated by the model

$$y_t + ay_{t-1} = e_t + ce_{t-1}$$

in which a, $|a| < 1$, and c are constants and $\{e_t\}$ is a white noise sequence with $var\{e_t\} = 1$.

a) Calculate the following covariances: $R_y(0)$, $R_y(1)$ and $R_y(2)$. [8]

b) Let $\hat{a}_N$ be the least squares estimate of a, based on the model

$$y_t + ay_{t-1} = \text{`noise'} \quad \text{for } t = 1, \ldots, N\,.$$

Calculate the asymptotic bias $a - \hat{a}_\infty$, where $\hat{a}_\infty = \lim_{N\to\infty} \hat{a}_N$. [8]

c) Now let

$$\hat{r}(1) = (1/N)\sum_{t=1}^{N} y_t y_{t-1} \quad \text{and} \quad \hat{r}(2) = (1/N)\sum_{t-1}^{N} y_t y_{t-2}$$

Show that the estimate

$$\hat{\hat{a}} = -\frac{\hat{r}(2)}{\hat{r}(1)}$$

is an unbiased estimate of a. [4]

You should assume throughout that sample covariances can be replaced by covariances as the sample size $N \to \infty$.

5. A 1-dimensional state-space structure considered in econometric modelling is

$$\begin{aligned} x_t &= -ax_{t-1} + bu_{t-1} \\ y_t &= x_t + e_t \\ z_t &= u_t + w_t \, . \end{aligned}$$

Here, $\{e_t\}$ and $\{w_t\}$ are independent, scalar, Gaussian white noise processes, with variances σ_x^2 and σ_u^2 respectively. $\{u_t\}$ is a deterministic signal.

Measurements are taken, not of the state x_t and the input u_t, but of the noise-corrupted state and control y_t and z_t, respectively.

Derive a difference equation model, relating the output y_t and the input z_t, of the form

$$A(z)y_t = B(z)z_{t-1} + C_1(z)e_t + C_2 w_{t-1} \, .$$

[2]

Take as data

$$\mathbf{y} = [y_1, \ldots, y_N]^T \quad \text{and} \quad \mathbf{z} = [z_o, \ldots, z_{N-1}]^T \, .$$

Construct the log likelihood function of a, b, σ_x^2 and σ_y^2, given $\mathbf{y}$ and $\mathbf{z}$. [12]

Hint: You should assume, for purposes of constructing this function, that initial terms can be ignored, i.e. you should set y_0, e_0 etc. to zero.

Now suppose that $\sigma_u^2 = 0$ and a and b are known. Show that the Maximum Likelihood estimate of σ_x^2 is

$$\hat{\sigma}_x^2 = (1/N)||A(a)\mathbf{y} - b\mathbf{u}||^2$$

where $A(a)$ is the matrix

$$\begin{bmatrix} 1 & 0 & . & . & 0 \\ a & 1 & . & . & 0 \\ 0 & a & 1 & . & 0 \\ . & . & . & . & . \\ 0 & . & 0 & a & 1 \end{bmatrix} .$$

[6]

Identification + Adaptive Control Exam 2010. Model Answers

1. (i) $x_{t+1} = Ax_t + bv_t$. So

$$E\{x_{t+1}x_{t+1}^T\} = E\{(Ax_t + bv_t)(Ax_t + bv_t)^T\}$$

Since x_t and e_t are independent, zero mean

$$E\{bv_t x_t^T b^T\} = 0. \text{ Hence}$$

$$E\{x_{t+1}x_{t+1}^T\} = E\{Ax_t x_t^T A^T\} + 0 + 0 + E\{bv_t v_t b^T\}$$

Hence

$$\underline{R_x(0) = AR_x(0) + \sigma^2 bb^T} \quad \text{(Lyapunov eqn.)}$$

We see that $E\{x_{t+k}x_t^T\} = AE\{x_{t+k-1}x_t^T\} + bE\{v_{t+k-1}x_t^T\}$

or $R_x(k) = AR_x(k-1) + 0$, provided $k \geq 1$

Hence,

$$\underline{R_x(k) = A^k R_x(0)} \qquad k \geq 1$$

Also $\underline{R_x(-k) = R_x(k)^T} = R_x(0)(A^T)^k$ for $k \geq 1$

And $R_y(k) = CR_x(k)C^T$ for all k

(ii) Write equations for y_t^1 and y_t^2 in state-space form

$$\begin{pmatrix} y_{t+1}^1 \\ y_{t+1}^2 \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -\alpha & 0 \end{bmatrix} \begin{bmatrix} y_t^1 \\ y_t^2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} v_t$$

Let $R_x(0) = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix}$. Then, from Lyapunov equation

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\alpha & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} \begin{bmatrix} 0 & -\alpha \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 \quad 1] \sigma^2$$

Hence $$\begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} = \begin{bmatrix} r_{22} & -\alpha r_{12} \\ -\alpha r_{12} & +\alpha^2 r_{11} \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Equating entries gives

$$r_{11} = r_{22} + \sigma^2, \quad r_{12} = -\alpha r_{21} + \sigma^2, \quad r_{22} = \alpha^2 r_{11} + \sigma^2$$

It follows

$$r_{11} = \alpha^2 r_{11} + 2\sigma^2, \quad \text{whence } r_{11} = \frac{2\sigma^2}{1-\alpha^2}$$

Then $r_{22} = \sigma^2\left[\frac{2\alpha^2}{1-\alpha^2} + 1\right] = \sigma^2\left[\frac{1+\alpha^2}{1-\alpha^2}\right]$

But $$\frac{z}{\varphi} = \frac{R_{y^2}(0)}{R_{y^1}(0)} = \frac{r_{22}}{r_{11}} = \frac{1+\alpha^2}{2}$$

It follows

$$1+\alpha^2 = 3/2. \quad \text{Hence } \alpha = \frac{1}{\sqrt{2}}$$

2(i) Write the model as $y = X\theta + e$ where $e = (e_1, e_2, \ldots, e_N)^T$ etc.

Note:

$$y - X\hat{\theta} = X\theta + e - X(X^TX)^{-1}X^T(X\theta + e) = (I - X(X^TX)^{-1}X^T)e. \text{ So}$$

$$E(y - X\hat{\theta})(y - X\hat{\theta})^T = \sigma^2(I - X(X^TX)^{-1}X^T)(I - X(X^TX)^{-1}X^T)$$

$$= \sigma^2(I - X(X^TX)^{-1}X^T).$$

Then $E(y - X\hat{\theta})^T(y - X\hat{\theta}) = E\,\mathrm{tr}\{(y - X\hat{\theta})^T(y - X\hat{\theta})\}$

$$= E\,\mathrm{tr}\{(y - X\hat{\theta})(y - X\hat{\theta})^T\} = E(\mathrm{tr}\, I_{N\times N} - (X^TX)^{-1}\mathrm{tr}\, XX^T)$$

$$= N - (X^TX)^{-1}\mathrm{tr}\{X^TX\} = N - \|X\|^{-1}\|X\| = \underline{N-1}$$

(ii) When d is unknown the appropriate LLS model is

$$y = [X \vdots \underline{1}]\begin{pmatrix}\theta\\ d\end{pmatrix} + e$$

in which $\underline{1} = (1, 1, \ldots, 1)^T$.

The standard formulae give

$$\begin{pmatrix}\hat{\theta}\\ \hat{d}\end{pmatrix} = \left(\begin{bmatrix}X^T\\ \underline{1}^T\end{bmatrix}[X : \underline{1}]\right)^{-1}\begin{bmatrix}X^Ty\\ 1^Ty\end{bmatrix} = \begin{bmatrix}X^TX & X^T\underline{1}\\ X^T1 & N\end{bmatrix}^{-1}\begin{bmatrix}X^Ty\\ 1^Ty\end{bmatrix}$$

$$= \cancel{N^{-2}}\left(\frac{1}{N}\sum_i x_i^2 - m_x^2\right)^{-1}\begin{bmatrix}1 & -m_x\\ -m_x & \frac{1}{N}\sum_i x_i^2\end{bmatrix}\begin{bmatrix}\frac{1}{N}\sum x_iy_i\\ m_y\end{bmatrix}\cancel{N^2}$$

$$= \left(\frac{1}{N}\sum_i x_i^2 - m_x^2\right)^{-1}\begin{bmatrix}\frac{1}{N}\sum_i x_iy_i - m_xm_y\\ \frac{1}{N}\sum_i x_i^2 m_y - \frac{1}{N}\sum x_iy_i)m_x\end{bmatrix}$$

$$= \frac{1}{\frac{1}{N}\sum_i(x_i - m_x)^2}\begin{bmatrix}\frac{1}{N}\sum_i(x_i - m_x)(y_i - m_y)\\ \frac{1}{N}\sum_i(x_i - m_x)^2 m_y - \sum_i(x_i - m_x)(y_i - m_x)m_x\end{bmatrix}$$

$$= \begin{bmatrix}\hat{r}_{xy}/\hat{r}_x\\ (\hat{r}_x m_y - r_{xy}m_x)/\hat{r}_x\end{bmatrix}$$

It follows $\hat{\theta} = \dfrac{\hat{r}_{xy}}{\hat{r}_x}$ and $\hat{d} = \dfrac{\hat{r}_x m_y - \hat{r}_{xy}m_x}{\hat{r}_x}$

Also

$$E\left[\begin{pmatrix}\theta - \hat{\theta}\\ d - \hat{d}\end{pmatrix}(-)^T\right] = \frac{1}{\hat{r}_x}\begin{bmatrix}\sigma^2/N & -m_x/N\\ -m_x/N & \sigma^2/N\,(\hat{r}_x + m_x^2)\end{bmatrix}$$

So $E[(\theta - \hat{\theta})^2] = \underline{\dfrac{1}{N}\dfrac{\sigma^2}{\hat{r}_x}}$ and $E[(d - \hat{d})^2] = \underline{\dfrac{\sigma^2}{N}\dfrac{(\hat{r}_x + m_x^2)}{\hat{r}_x}}$

3 (i) $\hat{\theta} = (x^T Q^{-1} x)^{-1} x^T Q^{-1} y$, $y = x\theta + e$. So
$\hat{\theta} = (x^T Q^{-1} x)^{-1} x^T Q^{-1} (x\theta + e) = \theta + (x^T Q^{-1} x)^{-1} x^T Q^{-1} e$,
and $E[\hat{\theta}] = \theta + (x^T Q^{-1} x)^{-1} x^T Q^{-1} E[e] = \theta + 0$. Unbiased

Log likelihood function is
$\ln p(y|\theta) = \text{const.} - \frac{1}{2}(y - x\theta)^T Q^{-1}(y - x\theta)$
$= \text{const.} - \frac{1}{2}\left[y^T Q^{-1} y - 2\theta x^T Q^{-1} y + \theta^2 x^T Q^{-1} x \right]$

It follows
$\partial^2/\partial\theta^2 \, p(y|\theta) = -x^T Q^{-1} x$. Hence $M = E_\theta\left[\frac{\partial^2}{\partial\theta^2} p(y|\theta)\right] = x^T Q^{-1} x$.

But $\text{var}\{\hat{\theta}\} = E\{|\hat{\theta} - \theta|^2\} = E\left|(x^T Q^{-1} x)^{-1} x^T Q^{-1} e\right|^2$
$= E\left[(x^T Q^{-1} x)^{-1}(x^T Q^{-1} x)(x^T Q^{-1} x)\right] = (x^T Q^{-1} x)^{-1}$.

We see $\text{var}\{\hat{\theta}\} = -M^{-1}$, i.e. $\hat{\theta}$ achieves CR lower bound.

$z = a_0\theta + v$. $\hat{z} = a_0\hat{\theta} = a_0 (x^T Q^{-1} x)^{-1} x^T Q^{-1}(x\theta + e)$
$= a_0\theta + a_0 (x^T Q^{-1} x)^{-1} x^T Q^{-1} e$

We see $E[|\hat{z} - z|^2] = E\left[\left|a_0 (x^T Q^{-1} x)^{-1} x^T Q^{-1} e - v\right|^2\right]$
$= a_0^2 (x^T Q^{-1} x)^{-1} + 0 + \sigma_0^2$.

It follows
$$\hat{z} - 2\sqrt{\frac{a_0^2}{x^T Q^{-1} x} + \sigma_0^2} \le z \le \hat{z} + 2\sqrt{\frac{a_0}{x^T Q^{-1} x} + \sigma_0^2} \quad \text{w.p. } .95.$$

(ii) Given scalar sequence $e_1, \ldots, e_N$, $N >> 1$, choose $k_{max} \in (1, N)$ and construct sample covariances
$$\hat{R}_{ee}(k) = (N - k_{max})^{-1} \Sigma^{!}_{k=k_{max}} (e_i - \bar{e})(e_{i-k} - \bar{e}), \quad k = 0, 1, \ldots, k_{max}$$
where $\bar{e} = (N - k_{max})^{-1} \Sigma^{!}_{i=k_{max}} e_i$

Now use the facts that, if $\{e_i\}$ is a white sequence
$E[e_i] = 0$,
$E[\hat{R}_{ee}(0)] = \sigma^2$ and $E[\hat{R}_{ee}(k)] = 0$ for $k \neq 0$
$E[\hat{R}_{ee}^2(k)] = \sigma^4/(N - k_{max})$ and
$E[\hat{R}_{ee}(k)\hat{R}_{ee}(l)] = 0$ for $k \neq l$

By law of large numbers, $\hat{R}_{ee}(0) \simeq \sigma^2$.
By central limit theorem, $\hat{R}_{ee}^2(k)$, $k = 1, 2, \ldots, k_{max}$ are also approx. normal.
They are therefore approx. independent.

These properties suggest test: $\bar{e} \simeq 0$ and
$-2 \cdot (N - k_{max})^{1/2} \le \hat{R}_{ee}(k)/\hat{R}_{ee}(0) \le +2(N - k_{max})^{1/2}$
for 95 % of k in $\{1, \ldots, k_{max}\}$

4 (i) $y_t + a y_{t-1} = e_t + c e_{t-1}$

$E[y_t y_{t-1} + a y_{t-1}^2] = E[e_t y_{t-1} + c e_{t-1} y_{t-1}] \Rightarrow R_y(1) + a R_y(0) = c R$

$E[y_t e_t + a y_{t-1} e_t] = E\{e_t^2\} + 0 \Rightarrow R_{ye}(0) = \sigma^2$

So $R_y(1) + a R_y(0) = c\sigma^2$ $\qquad = (1+c^2)$

$E[(y_t + a y_{t-1})^2] = E\{(e_t + c e_{t-1})^2\} \Rightarrow R_y(0) + 2a R_y(1) + a^2 R_y(0)$

Hence $R_y(0) + 2a(c\sigma^2 - a R_y(0)) + a^2 R_y(0) = (1+c^2)\sigma^2$ $\quad \sigma^2(1+c^2-2$

$\Rightarrow (1-a^2) R_y(0) = \sigma^2(1 + c^2 - 2ac)$. So $R_y(0) = \dfrac{\quad}{1-a^2}$

Then $R_y(1) = c\sigma^2 - a R_y(0) = \dfrac{(c - ca^2 - a - ac^2 + 2a^2c)\sigma^2}{(1-a^2)}$

So, $R_y(1) = \dfrac{(-a + c(1 + a^2 - ac))\sigma^2}{(1-a^2)}$

Also, $E[y_t y_{t-2}] + aE[y_t y_{t-2}] = E[(e_t + c e_{t-1}) y_{t-2}] = 0$. So

$R_y(2) + a R_y(1) = 0 \Rightarrow R_y(2) = \dfrac{-a(-a + c(1+a^2 - ac))\sigma^2}{(1-a^2)}$

(ii) Write
$$\overbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}^{y} = \overbrace{-a}^{\theta} \overbrace{\begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix}}^{X} + \text{"noise"}$$

Least squares estimate of a: $\hat{a}_N = -(X^T X)^{-1} X^T y = -\dfrac{(\frac{1}{N})\sum_{i=1}^{N} y_i y_{i-1}}{(\frac{1}{N})\sum_{i=1}^{N} y_i^2}$

By "ergodicity",
$$\hat{a}_N \to -\frac{R_y(1)}{R_y(0)} = \frac{a - c(1 + a^2 - ac)}{1 + c^2 - 2ac}$$

The bias is $a - \hat{a}_\infty = \dfrac{\cancel{a} + \cancel{ac^2} - 2a^2c - \cancel{a} + c + ca^2 - \cancel{ac^2}}{1 + c^2 - 2ac}$

$$= c(1 - a^2) / (1 + c(c - 2a))$$

(iii) $\hat{a}_N = -\hat{r}(2)/\hat{r}(1) = -\dfrac{(\frac{1}{N})\sum_{i=1}^{N} y_t y_{t-2}}{(\frac{1}{N})\sum_{i=1}^{N} y_t y_{t-1}}$

By ergodicity
$$\hat{a}_N \to -R_y(2)/R_y(1) = a \quad (\text{by part (i)}).$$

5. Eliminating x_t and u_t from system equations gives

$$y_t - e_t + a(y_{t-1} - e_{t-1}) = b(z_{t-1} - w_{t-1})$$

or

$$y_t + a y_{t-1} = b z_{t-1} - b w_{t-1} + e_t + a e_{t-1}$$

Ignoring initial conditions, we have

$$\underbrace{\begin{bmatrix} 1 & & 0 \\ a & 1 & \\ 0 & \ddots & \\ & a & 1 \end{bmatrix}}_{A(a)} \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{\underline{y}} = b \underbrace{\begin{bmatrix} z_0 \\ \vdots \\ z_{N-1} \end{bmatrix}}_{\underline{z}} - b\, \underline{W} + \begin{bmatrix} 1 & & 0 \\ a & \ddots & \\ 0 & a & 1 \end{bmatrix} \underline{e}$$

$\underline{W} = \begin{pmatrix} w_0 \\ \vdots \\ w_{N-1} \end{pmatrix}$, $\underline{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix}$

$$\Rightarrow \quad \underline{y} = b A^{-1}(a) \underline{z} - b A^{-1}(a) \underline{W} + \underline{e}$$

It follows that the likelihood function is

$$LF(\underline{y}, \underline{z}) = N\left(b A^{-1}(a) \underline{z},\ \sigma_x^2 I_{N\times N} + b^2 A^{-1}(a) A^{-T}(a) \sigma_w^2\right)$$

$$= \frac{1}{(2\pi)^{\frac{N}{2}} \det\left(\sigma_x^2 I_{N\times N} + b^2 A^{-1}(a) A^{-T}(a) \sigma_w^2\right)^{1/2}}$$

$$\times \exp\left\{ -\frac{1}{2} (\underline{y} - b A^{-1}(a) z)^T \left(\sigma_x^2 I + b^2 \sigma_w^2 A^{-1} A^{-T}\right)^{-1} (\underline{y} - b A^{-1}(a) z) \right\}$$

Since $\det(A(a))^{-1} = 1$, the normalization term is

$$1 / (2\pi)^{N/2} \det\left(\sigma_x^2 A(a) A^T(a) + b^2 I \sigma_w^2\right)^{1/2}$$

and $\exp\{ \ldots \} = \exp -\frac{1}{2} \varepsilon^T \left(\sigma_x^2 A(a) A^T(a) + b^2 I \sigma_w^2\right)^{-1} \varepsilon$

where

$$\varepsilon = A(a) \underline{y} - b \underline{z}$$

Hence

$$LLF(y, z) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log \det\left[\sigma_x^2 A(a) A^T(a) + b^2 I \sigma_w^2\right] - \frac{1}{2} \varepsilon^T \left(\sigma_x^2 A(a) A^T(a) + b^2 \sigma_w^2 I_{N\times N}\right)^{-1} \varepsilon$$

Now suppose $\sigma_w^2 = 0$ and a and b are known.

Then $\det(\sigma_x^2 A(a) A^T(a) + \ldots) = (\sigma_x^2)^N$. So

$$\frac{\partial}{\partial \sigma^2} LLF = -\frac{N}{2} \times \frac{1}{\sigma_x^2} + \frac{1}{2} \varepsilon^T [A(a) A^T a)]^{-1} \varepsilon \ \frac{1}{\sigma_x^4} = 0$$

Hence

$$\sigma_x^2 = \frac{\varepsilon^T [A(a) A^T(a)]^{-1} \varepsilon}{N},$$

where $\underline{\varepsilon} = A(a) \underline{y} - b \underline{z}$