STATISTICS: Problem Set 4-Solutions

October 11, 2007

1. Let X_1, X_2,..., X_T be iid with the following density $p_{X_i}(x_i) = \lambda e^{-\lambda x_i}$, $x_i > 0$, $\lambda > 0$.

(a) Find an expression for the log-likelihood function as a function of the parameter λ, conditioned on the available data.

The likelihood function is

$$L(\lambda | x_1, x_2, ..., x_T) := (\prod_{i=1}^{T} p_{X_i}(x_i)) = \lambda^T e^{-\lambda \sum_{i=1}^{T} x_i}.$$

So, the corresponding log-likelihood function is

$$LogL(\lambda | x_1, x_2, ..., x_T) := T ln\lambda - \lambda \sum_{i=1}^{T} x_i.$$

(b) Evaluate the score function $s_T(\lambda)$, that is the first derivative of the log-likelihood function.

$$s_T(\lambda) := \frac{d}{d\lambda} LogL(\lambda | x_1, x_2, ..., x_T) = \frac{T}{\lambda} - \sum_{i=1}^{T} x_i$$

(c) Work out the maximum likelihood estimator $\widehat{\lambda}_{ML}$

$$\widehat{\lambda}_{ML} := argmax_{\lambda} LogL(\lambda | x_1, x_2, ..., x_T)$$

The first-order condition is that the score function $s_T(\widehat{\lambda}_{ML})$ is equal to 0. This condition amounts to $\frac{T}{\lambda} = \sum_{i=1}^{T} x_i$. Therefore, we have:

$$\widehat{\lambda}_{ML} = \frac{T}{\sum_{i=1}^{T} x_i} = \frac{1}{\bar{x}}$$

(d) Try to work out its variance (Do not spend too much time on that question...)

In order to find the variance of $\widehat{\lambda}_{ML} = \frac{1}{\bar{x}}$, we can find the p.d.f. of the estimator and then simply compute its moments. Let's first determine the pdf of $X_1 + X_2 + ... + X_T$. This may be done by the change-of-variables method. Let

$$Y_1 = X_1 + X_2 + ... + X_T,$$
$$Y_2 = X_2 + ... + X_T,$$

$$Y_3 = X_3 + ... + X_T,$$
$$...$$
$$Y_T = X_T.$$

This means that

$$X_1 = Y_1 - Y_2,$$
$$X_2 = Y_2 - Y_3,$$
$$...$$
$$X_T = Y_T$$

and so, the Jacobian of the transformation is

$$J = \begin{bmatrix} 1 & -1 & 0 & 0 & ... & 0 \\ 0 & 1 & -1 & 0 & ... & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & ... & 1 \end{bmatrix}$$

The determinant of this Jacobian is 1: $\det(J) = 1$. Therefore, the joint density of $Y_1, .., Y_T$ is

$$p_{Y_1,..,Y_T}(y_1, .., y_T) = p_{X_1,..,X_T}(y_1 - y_2, .., y_T) \cdot 1 = \lambda^T e^{-\lambda y_1}, \quad 0 < y_T < y_{T-1} < ... < y_2 < y_1 < \infty.$$

The marginal density of y_1 is then

$$\begin{aligned} p_{Y_1}(y_1) &= \lambda^T e^{-\lambda y_1} \int_0^{y_1} dy_2 ... \int_0^{y_{T-3}} dy_{T-2} \int_0^{y_{T-2}} dy_{T-1} \int_0^{y_{T-1}} dy_T = \\ &= \lambda^T e^{-\lambda y_1} \int_0^{y_1} dy_2 ... \int_0^{y_{T-3}} dy_{T-2} \int_0^{y_{T-2}} y_{T-1} dy_{T-1} = \\ &= \lambda^T e^{-\lambda y_1} \int_0^{y_1} dy_2 ... \int_0^{y_{T-3}} \frac{y_{T-2}^2}{2} dy_{T-2} = ... = \lambda^T e^{-\lambda y_1} \frac{y_1^{T-1}}{(T-1)!} \\ \forall \ 0 &< y_1 < \infty \end{aligned}$$

This means that the distribution of the sum of the X_i's:

$$f_{X_1+X_2+...+X_T}(x) = \frac{\lambda^T}{(T-1)!} e^{-\lambda x} x^{T-1}$$

Now, using the rule of transformation once more, we have the distribution of the mean of X_i's:

$$f_{\bar{x}}(x) = f_{X_1+X_2+...+X_T}(Tx)|T| = \frac{\lambda^T T^T}{(T-1)!} x^{T-1} e^{-\lambda T x}$$

Finally, the third transformation leads to the pdf of the inverse of the mean:

$$f_{\frac{1}{\bar{x}}}(x) = f_{\bar{x}}(\frac{1}{x})| - \frac{1}{x^2}| = \frac{(\lambda T)^T}{(T-1)!} \left(\frac{1}{x}\right)^{T+1} e^{-\frac{\lambda T}{x}} = f_{\widehat{\lambda}_{ML}}$$

Now that we have the pdf of the estimator, we can compute its moments. Let's start with the expectation:

$$E(\widehat{\lambda}_{ML}) := \int_0^\infty x f_{\widehat{\lambda}_{ML}}(x) dx = \frac{1}{(T-1)!} \int_0^\infty \left(\frac{T\lambda}{x}\right)^T e^{-\frac{T\lambda}{x}} dx$$

Change the variable by taking $y = \frac{T\lambda}{x}$, $dy = -\frac{T\lambda}{x^2}dx$. Then

$$E(\widehat{\lambda}_{ML}) = \frac{T\lambda}{(T-1)!}\int_0^{\infty} y^{T-2}e^{-y}dy = \frac{T\lambda}{(T-1)!}(T-2)!$$

In the end, we have that:

$$E(\widehat{\lambda}_{ML}) = \frac{T\lambda}{(T-1)}$$

One can notice that the estimator is biased. In the very same fashion, one can compute the second moment: $E(\lambda^2_{ML}) = \frac{(T\lambda)^2}{(T-1)(T-2)}$ and finally get the variance of the estimator:

$$V(\widehat{\lambda}_{ML}) = \frac{(T\lambda)^2}{(T-1)(T-2)} - \left[\frac{T\lambda}{(T-1)}\right]^2 = \frac{(T\lambda)^2}{(T-1)^2(T-2)}.$$

(e) Evaluate the Information Matrix $i_T(\lambda)$.

$$i_T(\lambda) := -E(\frac{d^2}{d\lambda^2}LogL(\lambda|x_1, x_2, ..., x_T)) = -E(-\frac{T}{\lambda^2}) = \frac{T}{\lambda^2}$$

(f) Is $\widehat{\lambda}_{ML}$ an efficient estimator of λ? Comment.

First of all, we remarked that the estimator was a biased estimator of the parameter of interest. But, one can see that it is asymptotically unbiased and efficient since its variance reaches the CRLB.

2. Let X_1, X_2,..., X_T be iid with normal density $N(\mu, \sigma^2)$.

(a) Find an expression for the log-likelihood function as a function of the parameter $\theta = (\mu, \sigma^2)'$, conditioned on the available data.

The likelihood function is

$$L(\theta|x_1, x_2, ..., x_T) := \prod_{i=1}^{T} f_{X_i}(x_i) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^T \exp\left[-\sum_{i=1}^{T}\frac{(x_i-\mu)^2}{2\sigma^2}\right].$$

So, the log-likelihood function is

$$LogL(\theta|x_1, x_2, ..., x_T) := -T\ln\sqrt{2\pi} - \frac{T}{2}ln\sigma^2 - \sum_{i=1}^{T}\frac{(x_i-\mu)^2}{2\sigma^2}.$$

(b) Evaluate the score function $s_T(\theta)$.

$$s_T(\theta) = \begin{pmatrix} \frac{\partial}{\partial\mu}LogL(\theta|x_1, ..., x_T) \\ \frac{\partial}{\partial\sigma^2}LogL(\theta|x_1, ..., x_T) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{T}\frac{(x_i-\mu)}{\sigma^2} \\ -\frac{T}{2\sigma^2} + \sum_{i=1}^{T}\frac{(x_i-\mu)^2}{2\sigma^4} \end{pmatrix}$$

(c) Work out the maximum likelihood estimator $\widehat{\theta}_{ML}$.

Taking the score being equal to zero, we obtain:

$$\hat{\theta}_{ML} = \begin{pmatrix} \hat{\mu}_{ML} \\ \hat{\sigma}^2_{ML} \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \sum_{i=1}^{T}\frac{(x_i-\bar{x})^2}{T} \end{pmatrix}$$

(d) Work out its variance.

$\hat{\mu}_{ML} = \bar{x}$

$$E(\bar{x}) = E(\sum_{i=1}^{T} \frac{x_i}{T}) = \frac{1}{T}\sum_{i=1}^{T} E(x_i) = \mu$$
$$V(\bar{x}) = V(\sum_{i=1}^{T} \frac{x_i}{T}) = \frac{1}{T^2}\sum_{i=1}^{T} V(x_i) = \frac{\sigma^2}{T}$$

$\hat{\sigma}^2_{ML}$ Note that since X_1, X_2,..., X_T are iid with normal density $N(\mu, \sigma^2)$, $\frac{\sum_{i=1}^{T}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{T-1}$.

$$E(\hat{\sigma}^2_{ML}) = E\left(\sum_{i=1}^{T} \frac{(x_i-\bar{x})^2}{T}\right) = \frac{\sigma^2}{T}E\left(\frac{\sum_{i=1}^{T}(x_i-\bar{x})^2}{\sigma^2}\right) = \frac{\sigma^2}{T}E(\chi^2_{(T-1)}) = \sigma^2\frac{T-1}{T}$$
$$V(\hat{\sigma}^2_{ML}) = V\left(\sum_{i=1}^{T} \frac{(x_i-\bar{x})^2}{T}\right) = \frac{\sigma^4}{T^2}V\left(\frac{\sum_{i=1}^{T}(x_i-\bar{x})^2}{\sigma^2}\right) = \frac{\sigma^4}{T^2}V(\chi^2_{(T-1)}) = \sigma^4\frac{2(T-1)}{T^2}$$

(e) Evaluate the Information Matrix $i_T(\theta)$.

$$i_T(\theta) = -E\begin{pmatrix} \frac{\partial^2}{\partial\mu^2}LogL & \frac{\partial^2}{\partial\mu\partial\sigma^2}LogL \\ \frac{\partial^2}{\partial\mu\partial\sigma^2}LogL & \frac{\partial^2}{\partial(\sigma^2)^2}LogL \end{pmatrix}$$

Let $Q(\theta)$ be the matrix of second derivatives, we have:

$$Q(\theta) = \begin{pmatrix} -\frac{T}{\sigma^2} & , -\frac{\sum_{i=1}^{T}(x_i-\mu)}{\sigma^4} \\ -\frac{\sum_{i=1}^{T}(x_i-\mu)}{\sigma^4} & , \frac{T}{2\sigma^4} - \frac{\sum_{i=1}^{T}(x_i-\mu)^2}{\sigma^6} \end{pmatrix}$$

Now, taking minus the expectation of Q, we have:

$$i(\theta) = -E(Q(\theta)) = \begin{pmatrix} \frac{T}{\sigma^2} & 0 \\ 0 & \frac{T}{2\sigma^4} \end{pmatrix}$$

(f) Is the maximum likelihood estimator of σ^2 an unbiased estimator of σ^2? Does is achieve the Cramer-Rao lower bound?

The CRLB is the inverse of the information matrix, i.e. :

$$(i_T(\theta))^{-1} = \begin{pmatrix} \frac{\sigma^2}{T} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{pmatrix}$$

First of all, one can see that only the MLE of μ is unbiased. Therefore, efficiency in finite sample can only be considered for it, but not for σ^2_{ML} since it is biased. Having a look at the inverse of the information matrix, one can see that the variance of $\bar{x}$ is equal to the CRLB. Therefore, $\bar{x}$ is an efficient estimator of the parameter μ.
Now, if we consider the asymptotic properties of both estimator, one can see that they are both unbiased, and also efficient. This is a general property of ML estimators.

3.
(a) Define a maximum likelihood estimator.
A maximum likelihood estimator is a function of the observations which maximizes the likelihood function. Formally, $\lambda_{ML} = argmax_{\lambda}LogL(\lambda)$. That is why the necessary condition is the score to be equal to 0.

(b) Let X_1, X_2,..., X_T be iid with Poisson distribution given by:

$$f_{X_i}(x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Derive the log-likelihood function for the sample, and compute the *Cramer-Rao* lower bound.

$$L(\lambda|x_1,...,x_T) = \prod_{i=1}^{T} f_{X_i}(x_i) = \frac{e^{-\lambda T}\lambda^{\sum X_i}}{\Pi(x_i!)}$$
$$LogL(\lambda|x_1,...,x_T) = -\lambda T + \sum_{i=1}^{T} x_i ln(\lambda) - \sum_{i=1}^{T} ln(x_i!)$$
$$s_T(\lambda) := \frac{\partial LogL(\lambda|x_1,...,x_T)}{\partial\lambda} = -T + \frac{\sum_{i=1} x_i}{\lambda}$$
$$CRLB := (-E(\frac{\partial s_T(\lambda)}{\partial\lambda}))^{-1} = (-E\left(-\sum_{i=1}\frac{x_i}{\lambda^2}\right))^{-1} = \left(\frac{T\lambda}{\lambda^2}\right)^{-1} = \frac{\lambda}{T}$$

(c) Derive the maximum likelihood estimator for λ.

By definition, the MLE is such that the score is equal to 0, which means in our case that $\hat{\lambda}_{ML} = \overline{x}$.

(d) Derive the mean and variance of the ML estimator and show whether or not its variance achieves the Cramer-Rao bound.

$$E(\overline{x}) = \lambda$$
$$Var(\overline{x}) = \frac{\lambda}{T}$$

We see that our MLE is unbiased and that its sample variance is equal to the CRLB. Therefore, our estimator is efficient.

4. Consider the process:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \tilde{} IN_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{matrix} \sigma^2 & \varphi \\ \varphi & \omega^2 \end{matrix} \right] \text{ for } t = 1, 2..., T.$$

(a) You want to estimate the parameter β in a conditional linear model relating y_t to z_t. How can this be justified? Derive β as a function of the parameters of the process above.

Regression of y_t on z_t is justified because y_t and z_t are correlated.

Recall that generally if $\begin{pmatrix} y_t \\ z_t \end{pmatrix} \tilde{} IN_2 \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{matrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{matrix} \right]$, $t \in 1,..,T$, then we know the conditional distribution of y_t/z_t:

$$y_t/z_t \sim N_1 \left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(z_t - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \right).$$

denote by $u_t = y_t - E(y_t/z_t)$. Then

$$y_t = \left(\mu_1 - \mu_2 \frac{\sigma_{12}}{\sigma_2^2} \right) + \frac{\sigma_{12}}{\sigma_2^2} z_t + u_t.$$

So, regressing y_t on z_t gives us

$$\alpha = \mu_1 - \mu_2 \frac{\sigma_{12}}{\sigma_2^2},$$
$$\beta = \frac{\sigma_{12}}{\sigma_2^2}.$$

In our problem, $\mu_1 = \mu_2 = 0$, $\sigma_{12} = \varphi$, $\sigma_2^2 = \omega^2$. Therefore,

$$\alpha = 0,$$
$$\beta = \frac{\varphi}{\omega^2}.$$

(b) Write down the likelihood function for the bivariate process (use the assumption of independence) and find the maximum likelihood estimator $\widehat{\beta}$ of β.

Let $\underline{\mathrm{x}}_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}$. Then using the formula for the $n-$variate normal density function

$$f_{\underline{x}}(\underline{x}) = (2\pi)^{-n/2}|\Sigma|^{-1/2}exp(-\tfrac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})),$$

we know the density function of $\underline{x}_t$:

$$f(\underline{x}_t) = (2\pi)^{-1}(\sigma^2\omega^2-\varphi^2)^{-1/2}exp(-\frac{1}{2}\underline{x}_t'\Sigma^{-1}\underline{x}_t) = \; = (2\pi)^{-1}(\omega^2(\sigma^2-\frac{\varphi^2}{\omega^2}))^{-1/2}exp(-\frac{1}{2}\underline{x}_t'\Sigma^{-1}\underline{x}_t).$$

Alternatively, one may think of density function for $\underline{x}_t$ as $f_{y_t/z_t} * f_{z_t}$.

Let us define $\sigma_{1.2}^2 = \sigma^2 - \frac{\varphi^2}{\omega^2}$ (it is the conditional variance of y_t/z_t). Then the likelihood function for the bivariate process $\underline{x}_t$ is

$$L = \prod_{t=1}^{T} f(\underline{x}_t) = (2\pi)^{-T}(\omega^2\sigma_{1.2}^2)^{-T/2}\prod_{t=1}^{T} exp(-\frac{1}{2}\underline{x}_t'\Sigma^{-1}\underline{x}_t).$$

Taking logs gives

$$LogL = -Tlog(2\pi) - \frac{T}{2}log(\omega^2) - \frac{T}{2}log(\sigma_{1.2}^2) - \frac{1}{2}\sum_{t=1}^{T}\underline{x}_t'\Sigma^{-1}\underline{x}_t.$$

$$\begin{aligned}
\underline{x}_t'\Sigma^{-1}\underline{x}_t &= (y_t \quad z_t)\frac{1}{\sigma^2\omega^2-\varphi^2}\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{bmatrix}\begin{pmatrix}y_t\\ z_t\end{pmatrix} = \\
&= \frac{1}{\sigma^2\omega^2-\varphi^2}(\omega^2y_t^2 - 2\varphi y_t z_t + \sigma^2 z_t^2) = \\
&= \frac{1}{\sigma_{1.2}^2}\left(y_t^2 - 2\frac{\varphi}{\omega^2}y_t z_t + \frac{\sigma^2}{\omega^2}z_t^2\right) = \\
&= \frac{1}{\sigma_{1.2}^2}\left[\left(y_t - \frac{\varphi}{\omega^2}z_t\right)^2 + \left(\frac{\sigma^2}{\omega^2} - \frac{\varphi^2}{\omega^4}\right)z_t^2\right] = \\
&= \frac{1}{\sigma_{1.2}^2}\left[\left(y_t - \frac{\varphi}{\omega^2}z_t\right)^2 + \frac{1}{\omega^2}\left(\sigma^2 - \frac{\varphi^2}{\omega^2}\right)z_t^2\right] = \\
&= \frac{1}{\sigma_{1.2}^2}\left(y_t - \frac{\varphi}{\omega^2}z_t\right)^2 + \frac{1}{\omega^2}z_t^2.
\end{aligned}$$

Thus

$$LogL = -Tlog(2\pi) - \frac{T}{2}log(\omega^2) - \frac{T}{2}log(\sigma_{1.2}^2) - \frac{1}{2\sigma_{1.2}^2}\sum_{t=1}^{T}\left(y_t - \frac{\varphi}{\omega^2}z_t\right)^2 - \frac{1}{2\omega^2}\sum_{t=1}^{T}z_t^2.$$

The likelihood problem now is to maximize $logL$ with respect to $\frac{\varphi}{\omega^2}$, $\sigma_{1.2}^2$ and ω^2. Note that the parameters are variation free (in the sense that fixing ω^2 for example does not restrict the values that the remaining parameters can take on and similarly, for the other parameters).

From part (a) we know that $\frac{\varphi}{\omega^2} = \beta$. Then we have

$$LogL = -Tlog(2\pi) - \frac{T}{2}log(\omega^2) - \frac{T}{2}log(\sigma_{1.2}^2) - \frac{1}{2\sigma_{1.2}^2}\sum_{t=1}^{T}(y_t - \beta z_t)^2 - \frac{1}{2\omega^2}\sum_{t=1}^{T}z_t^2.$$

Differentiate this function with respect to β and equate the derivative with zero:

$$\frac{1}{\sigma_{1.2}^2}\sum_{t=1}^{T}z_t\,(y_t - \beta z_t) = 0.$$

Hence,

$$\hat{\beta} = \frac{\sum_{t=1}^{T} z_t y_t}{\sum_{t=1}^{T} z_t^2}.$$

This is a ML estimate of β.

The ML estimates of the other parameters can be found in a similar way. Differentiating $LogL$ with respect to $\sigma_{1.2}^2$ and ω^2 we obtain

$$\begin{aligned} \hat{\sigma}_{1.2}^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \beta z_t)^2, \\ \hat{\omega}^2 &= \frac{1}{T}\sum_{t=1}^{T} z_t^2. \end{aligned}$$

Since $\hat{\beta} = \frac{\hat{\varphi}}{\hat{\omega}^2}$, once we know $\hat{\omega}^2$, we also have an estimate for $\hat{\varphi}$:

$$\hat{\varphi} = \hat{\beta}\hat{\omega}^2 = \frac{\sum_{t=1}^{T} z_t y_t}{\sum_{t=1}^{T} z_t^2}\frac{\sum_{t=1}^{T} z_t^2}{T} = \frac{\sum_{t=1}^{T} z_t y_t}{T}.$$

Besides, since $\hat{\sigma}_{1.2}^2 = \hat{\sigma}^2 - \frac{\hat{\varphi}^2}{\hat{\omega}^2}$, and we know the estimates for $\sigma_{1.2}^2$, φ, and $\frac{\varphi}{\omega^2}$, we can also estimate $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\beta} z_t\right)^2 + \hat{\beta}\frac{\sum_{t=1}^{T} z_t y_t}{T}.$$

5. Let

$$y_t = \beta z_t + u_t,$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

with $|\rho| < 1$ and $\varepsilon_t \tilde{\ } IN(0,1)$.

You may also assume that the z_t are determined outside the model, and can be taken as fixed in repeated samples.

Obtain the score and information matrix for the problem and show that the maximum likelihood estimators of β and ρ are independent of each other (i.e. the information matrix is diagonal).

Briefly describe how you would calculate the estimators.

Given the AR(1) process for u_t, the process for y_t can be represented in the form

$$y_t = \beta z_t + \rho u_{t-1} + \varepsilon_t.$$

$\varepsilon_t \tilde{\ } IN(0,1)$ means that $(y_t - \beta z_t - \rho u_{t-1})\tilde{\ } IN(0,1)$ or $(y_t - \beta z_t - \rho(y_{t-1} - \beta z_{t-1}))\tilde{\ } IN(0,1)$ and so, the likelihood function for this process is

$$\begin{aligned} L(\beta, \rho | \underline{y}_t, \underline{z}_t) &= \prod_{t=1}^{T}\left[\frac{1}{\sqrt{2\pi}} exp(-\frac{(y_t - \beta z_t - \rho(y_{t-1} - \beta z_{t-1}))^2}{2})\right] = \\ &= \frac{1}{(2\pi)^{T/2}} exp\left(-1/2\sum_{t=1}^{T}(y_t - \beta z_t - \rho(y_{t-1} - \beta z_{t-1}))^2\right). \end{aligned}$$

The corresponding log-likelihood function is

$$logL(\beta,\rho|\underline{y}_t,\underline{z}_t) = -\frac{T}{2}log(2\pi) - 1/2\sum_{t=1}^{T}(y_t - \beta z_t - \rho y_{t-1} + \rho\beta z_{t-1})^2.$$

The score vector $q_T(\beta,\rho)$ is defined as a vector of the first derivatives of $logL$ with respect to β and ρ:

$$q_T(\beta,\rho) = \begin{pmatrix} \frac{\partial}{\partial\beta}LogL \\ \frac{\partial}{\partial\rho}LogL \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T}(y_t - \beta z_t - \rho y_{t-1} + \rho\beta z_{t-1})(z_t - \rho z_{t-1}) \\ \sum_{t=1}^{T}(y_t - \beta z_t - \rho y_{t-1} + \rho\beta z_{t-1})(y_{t-1} - \beta z_{t-1}) \end{pmatrix}$$

The information matrix $i_T(\beta,\rho)$ is "-" expectation of the matrix $Q(\beta,\rho)$ of the second partial derivatives of $logL$ with respect to β and ρ: $i_T(\beta,\rho) = -E(Q(\beta,\rho))$, where

$$Q(\beta,\rho) = \begin{pmatrix} \frac{\partial^2}{\partial\beta^2}LogL & \frac{\partial^2}{\partial\beta\partial\rho}LogL \\ \frac{\partial^2}{\partial\beta\partial\rho}LogL & \frac{\partial^2}{\partial\rho^2}LogL \end{pmatrix} =$$

$$= \begin{pmatrix} -\sum_{t=1}^{T}(\rho z_{t-1} - z_t)^2, & \sum_{t=1}^{T}((z_t - \rho z_{t-1})(\beta z_{t-1} - y_{t-1}) - z_{t-1}(y_t - \beta z_t - \rho y_{t-1} + \rho\beta z_{t-1})) \\ \sum_{t=1}^{T}((z_t - \rho z_{t-1})(\beta z_{t-1} - y_{t-1}) - z_{t-1}(y_t - \beta z_t - \rho y_{t-1} + \rho\beta z_{t-1})), & -\sum_{t=1}^{T}(\beta z_{t-1} - y_{t-1})^2 \end{pmatrix}$$

The term outside the diagonal is

$$\begin{aligned} \beta z_t z_{t-1} - z_t y_{t-1} - \rho\beta z_{t-1}^2 + \rho z_{t-1}y_{t-1} - z_{t-1}y_t + \beta z_{t-1}z_t + \rho z_{t-1}y_{t-1} - \rho\beta z_{t-1}^2 = \\ = 2\beta z_t z_{t-1} - 2\rho\beta z_{t-1}^2 + (2\rho z_{t-1} - z_t)y_{t-1} - z_{t-1}y_t = \\ = 2\beta z_t z_{t-1} - 2\rho\beta z_{t-1}^2 + (2\rho z_{t-1} - z_t)(\beta z_{t-1} + u_{t-1}) - z_{t-1}(\beta z_t + u_t) = \\ = (2\rho - 1)z_{t-1}u_t - z_t u_{t-1}. \end{aligned}$$

Since z_t are taken as fixed in repeated samples, z_t is uncorrelated with the leads and lags of y_t. So,

$$E(z_{t-1}u_t) = E(z_{t-1})E(u_t) = 0,$$
$$E(z_t u_{t-1}) = E(z_t)E(u_{t-1}) = 0.$$

Therefore, we get:

$$i_T(\beta,\rho) = \begin{pmatrix} \sum_{t=1}^{T}(\rho z_{t-1} - z_t)^2, & 0 \\ 0, & \frac{T}{1-\rho^2} \end{pmatrix}$$

Note, that if not the condition that z_t are taken as fixed in repeated samples, i_T would not necessarily be diagonal and $\hat{\beta}$ and $\hat{\rho}$ would then be not independent of each other.

The ML estimators of β and ρ turn the score vector to zero:

$$\begin{cases} \sum_{t=1}^{T}(y_t - \hat{\beta}(z_t - \hat{\rho}z_{t-1}) - \hat{\rho}y_{t-1})(z_t - \hat{\rho}z_{t-1}) = 0 \\ \sum_{t=1}^{T}(y_t - \hat{\beta}z_t - \hat{\rho}(y_{t-1} - \hat{\beta}z_{t-1})(y_{t-1} - \hat{\beta}z_{t-1}) = 0 \end{cases}$$

$$\hat{\beta} = \frac{\sum_{t=1}^{T}(y_t - \hat{\rho}y_{t-1})(z_t - \hat{\rho}z_{t-1})}{\sum_{t=1}^{T}(z_t - \hat{\rho}z_{t-1})^2} \quad (1)$$

$$\hat{\rho} = \frac{\sum_{t=1}^{T}(y_t - \hat{\beta}z_t)(y_{t-1} - \hat{\beta}z_{t-1})}{\sum_{t=1}^{T}(y_{t-1} - \hat{\beta}z_{t-1})^2} = \frac{\sum_{t=1}^{T}\hat{u}_t\hat{u_{t-1}}}{\sum_{t=1}^{T}\hat{u_{t-1}}^2} \quad (2)$$

To calculate the estimators, one should first regress y_t on z_t and obtain the residuals $\{\hat{u}_t\}_{t=1}^{T}$ of this regression. Then $\hat{\rho}$ is calculated according to (2) and after that, given $\hat{\rho}$, $\hat{\beta}$ is found from (1).