92.3.5. *Efficiency of maximum likelihood*—Solutions.[1] Two solutions have been proposed independently by Peter C.B. Phillips (the poser of the problem) and Benedikt M. Pötscher. These solutions, which are published below, contain a different interesting proof of the consistency of the estimator.

Solution Proposed by Peter C.B. Phillips

(i) The log likelihood of the model

$$y_t = bx_t + u_t, \quad u_t \equiv \text{i.i.d. } N(0, b^2) \quad (t = 1, \ldots, n) \tag{1}$$

is given by

$$L_n(b) = -(n/2)\ln 2\pi - (n/2)\ln b^2 - (1/2b^2)\sum_1^n (y_t - bx_t)^2.$$

Let b_0 be the true value of b in (1). Note that $L_n(b)$ is a continuous function of b on the union of the two half lines $(-\infty, 0) \cup (0, \infty)$ and $L_n(b) \to -\infty$ as $b \to 0, \pm\infty$. Hence, $L_n(b)$ achieves a global maximum at some finite value of $b \neq 0$, say $\tilde{b}$. Hence

$$n^{-1}L_n(\tilde{b}) \geq n^{-1}L_n(b_0). \tag{2}$$

Let $\hat{b}$ be the OLS estimator of b in (1). Since $\sum_1^n x_t^2 \to \infty$, we have $\hat{b} \to b_0$ a.s. and $n^{-1}\sum_1^n \hat{u}_t^2 \to E(u_t^2) = b_0^2$ a.s. where $\hat{u}_t = y_t - \hat{b}x_t$ are the OLS residuals. Now

$$\begin{aligned} n^{-1}L_n(b) &= -(\tfrac{1}{2})\ln 2\pi - (\tfrac{1}{2})\ln b^2 \\ &\quad - (1/2b^2)\left\{n^{-1}\sum_1^n \hat{u}_t^2 + (b - \hat{b})^2 n^{-1}\sum_1^n x_t^2\right\} \\ &\to -(\tfrac{1}{2})\ln 2\pi - (\tfrac{1}{2})\ln b^2 - (1/2b^2)\{b_0^2 + (b - b_0)^2 m_x\} \text{ a.s.} \\ &= L(b), \text{ say.} \end{aligned}$$

The convergence is also uniform in b, since the convergences $n^{-1}\sum \hat{u}_t^2 \to_{\text{a.s.}} b_0^2$, $\hat{b} \to_{\text{a.s.}} b_0$ and $n^{-1}\Sigma_1^n x_t^2 \to m_x$ are all independent of b.

We may write

$$\begin{aligned} L(b) &= \{-(\tfrac{1}{2})\ln 2\pi - (\tfrac{1}{2})\ln b^2 - b_0^2/2b^2\} + \{-(1/2b^2)(b - b_0)^2 m_x\} \\ &= g(b) + h(b), \text{ say.} \end{aligned}$$

The function $h(b)$ has a global maximum at $b = b_0$, while $g(b)$ has global maxima at $b = \pm b_0$. Thus, the global maximum of $L(b)$ occurs at $b = b_0$, and this optimum value is unique. In view of the inequality (2) and the uniform convergence of $n^{-1}L_n(b)$, we deduce that $\tilde{b} \to b_0$ a.s.

The asymptotic distribution of $\tilde{b}$ is given by the limit

$$\sqrt{n}(\tilde{b} - b_0) \rightarrow_d N(0, v^2(b_0)),$$

where

$$v^2(b_0) = \left\{ \lim_{n\to\infty} E\left(-\frac{1}{n} \frac{d^2 L_n(b_0)}{db^2} \right) \right\}^{-1}.$$

Now

$$n^{-1} dL_n(b)/db = (-1/b) + (1/b^3) n^{-1} \sum_1^n (y_t - bx_t)^2 + (1/b^2) n^{-1} \sum_1^n (y_t - bx_t) x_t$$

and

$$n^{-1} d^2 L_n(b)/db^2 = 1/b^2 - (3/b^4) n^{-1} \sum_1^n (y_t - bx_t)^2 - (4/b^3) n^{-1} \sum_1^n (y_t - bx_t) x_t - (1/b^2) n^{-1} \sum_1^n x_t^2.$$

Hence,

$$n^{-1} E(d^2 L_n(b_0)/db^2) = 1/b_0^2 - (3/b_0^4) b_2^0 - (1/b_0^2) n^{-1} \sum_1^n x_t^2$$

and

$$\lim_{n\to\infty} E\left(-\frac{1}{n} \frac{d^2 L_n(b_0)}{db^2} \right) = \frac{2 + m_x}{b_0^2}.$$

Thus,

$$v^2(b_0) = b_0^2/(2 + m_x).$$

(ii) By contrast, if $\hat{b}$ is the OLS estimator of b, we have

$$\sqrt{n}(\hat{b} - b_0) \rightarrow_d N(0, b_0^2/m_x).$$

Since $v^2(b_0) < b_0^2/m_x$, it follows that the MLE $\tilde{b}$ is asymptotically more efficient than the OLS estimator $\hat{b}$.