1) \( y_t = g(x_t) + \epsilon_t \) \( \epsilon_t \sim N(0, \sigma^2) \) \( x_t \) has density \( f_x \).

We (falseely) fit the model \( y_t = \beta x_t + e_t \) \( \epsilon_t \sim N(0, \sigma^2) \) to no data.

(c) \( \hat{\beta} = \frac{\sum y_t x_t}{\sum x_t^2} \)

(vi) By the continuous mapping theorem we have

\[
\hat{\beta} = \frac{\frac{1}{T} \sum y_t x_t}{\frac{1}{T} \sum x_t^2} \rightarrow \mathbb{E}(y_t x_t) = \frac{\int g(x) x f(x) dx}{\int g(x) f(x) dx} = \beta_0.
\]

(vii) There are 2 ways to approach this problem.

Either consider the marginal distribution of \( y_t \) for both the correct model and the best fitting linear model or consider the joint distribution of \( (y_t, x_t) \) for both the best fitting and correct model.

In the case of the marginal distributions the true density is:

\[
h(y) = \int \left[ y - g(x) \right] f_x(x) dx.
\]

For the correct model it is:

\[
\hat{f}_p(y) = \int \left[ y - \beta_0 x \right] f_x(x) dx.
\]
The KL-criterion is then $E_n \left[ \log \frac{\hat{p}(y)}{n(y)} \right] = \phi \alpha$ at does not give an interesting explicit expression.

Let us, instead, consider the joint distributions.

The true joint distribution is

$$p(y|x) f_x(x) = \mathcal{N}(y - g(x),\sigma_x^2) f_x(x)$$

The best fitting distribution is

$$f_y|x(y) f_x(x) = \mathcal{N}(y - \beta_0 x) f_x(x).$$

Thus the KL-criterion is

$$E \left[ \log \frac{\mathcal{N}(y - \beta x, \sigma_x^2) f_x(x)}{\mathcal{N}(y - g(x), \sigma^2) f_x(x)} \right]$$

$$= E \left\{ -\frac{1}{2 \sigma^2} (y - \beta x)^2 + \frac{1}{2 \sigma^2} (y - g(x))^2 \right\}$$

$$= -\frac{1}{2 \sigma^2} \int (\beta x - g(x))^2 \delta f(x) \, dx$$

[after some calculation and using mean $E(Y|X) = E(g(X)|X)$]

This has a rather nice form. It will always be negative (as theory tells us), but the better the approximation the closer it will be to zero.
2) Let us suppose that $Y_t$ is a discrete r.v. and $X_t$ are regressors which we know to influence $Y_t$. If we treat $X_t$ as random then conditional on $X_t$ we have

$E(Y_t | X_t) = e^{\beta X_t}$ and $\text{Var}(Y_t | X_t) = e^{2\beta X_t} \left[ 1 + \xi e^{\beta X_t} \right]$, where $\xi \geq 0$. The estimating equation

$\Theta(\beta) = \sum_t (Y_t - e^{\beta X_t}) X_t = 0$ is used to estimate $\beta$.

We denote this estimate as $\hat{\beta}$.

By the delta method we have

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_T).$$

(a) $V_T = A_T^{-1} B_T A_T^{-1}$, where

$A_T = \mathbb{E} \left\{ \frac{1}{T} \sum_t e^{\beta X_t} X_t^2 \right\} = \mathbb{E} \left\{ e^{\beta X_t} X_t^2 \right\}$

$B_T = \text{Var} \left\{ \frac{1}{T} \sum_t (Y_t - e^{\beta X_t}) X_t \right\}$

$$= \frac{1}{T} \sum_t \mathbb{E} \left\{ \text{Var}(Y_t - e^{\beta X_t} | X_t) X_t^2 \right\}$$

$$= \frac{1}{T} \sum_t \mathbb{E} \left\{ X_t^2 e^{\beta X_t} \left[ 1 + \frac{\xi}{T} e^{\beta X_t} \right] \right\}$$

$$= \mathbb{E} \left[ X_t^2 e^{\beta X_t} \right] + \frac{\xi}{T} \mathbb{E} \left[ X_t^2 e^{2\beta X_t} \right]$$

(ii) To derive explicit expressions for $A_T$ and $B_T$ when $X_t$ are standard normal distribution, we use the
\[
\frac{1}{\sqrt{2\pi}} \int x e^{-\frac{1}{2} x^2} \, dx = \frac{e^{\sigma^2/2}}{\sqrt{\pi}} \int x e^{-\frac{1}{2} (x-\bar{x})^2} \, dx \\
= \frac{e^{\sigma^2/2}}{\sqrt{\pi}} \left[ \frac{\sqrt{\pi} (x-\bar{x}) e^{-\frac{1}{2} (x-\bar{x})^2}}{\sqrt{\pi}} \right] + \frac{\sigma}{\sqrt{2\pi}} \int e^{-\frac{1}{2} (x-\bar{x})^2} \, dx \\
= \frac{e^{\sigma^2/2}}{\sqrt{\pi}} \left[ \sigma + \sigma \right] = \sigma e^{\sigma^2/2} 
\]

(1)

\[
\frac{1}{\sqrt{2\pi}} \int x^2 e^{-\frac{1}{2} x^2} \, dx = \\
e^{\sigma^2/2} \left[ \frac{\sqrt{\pi} x e^{-\frac{1}{2} (x-\bar{x})^2}}{\sqrt{\pi}} \right] + \frac{2\sigma}{\sqrt{2\pi}} \int x e^{-\frac{1}{2} (x-\bar{x})^2} \, dx - \frac{\sigma^2}{\sqrt{\pi}} \int e^{-\frac{1}{2} (x-\bar{x})^2} \, dx \\
= e^{\sigma^2/2} \left[ 1 + \sigma^2 \right] 
\]

(2)

Thus

\[
A_T = \int e^{B x_t} x_t e(x) \, dx = e^{B^2/2} (\beta + 1) 
\]

(3)

(\text{iii}) \text{ Estimate } B_T \text{ with } \frac{1}{T} \sum_{t=1}^{T} \left( y_t - e^{B x_t} \right) x_t^2 \overset{\hat{\beta}}{=} B
\[ \hat{B} = e^{\hat{B}^2 / 2} (1 + \hat{B}^4) + \varepsilon e^{(2\hat{B})^2 / 2} (1 + (2\hat{B})^2) \]

Now we can solve for \( \varepsilon \), to obtain an estimate of \( \varepsilon \).

There also exists graphical methods for estimating \( \varepsilon \) too.
2) we have the survival likelihood:

\[ L_t(\theta) = \sum_i S_i \log f(T_i; \theta) + \sum_i (1-S_i) \log F(t_i; \theta). \]

(a) since the observations are not taking expectations give

\[ \mathbb{E} \{ L_t(\theta) \} = \pi \mathbb{E} \{ \log f(T_i; \theta) \} + (1-\pi) \mathbb{E} \{ \log F(t_i; \theta) \} \]

since \( S_i \) and \( (T_i, \xi_i) \) are independent.

Differentiating \( \mathbb{E} \{ \log f(T_i; \theta) \} \) with \( \theta \) it is easily seen that \( \frac{\partial}{\partial \theta} \mathbb{E} \{ \log f(T_i; \theta) \} = 0 \) at \( \theta = \theta_0 \).

We now consider the term \( \mathbb{E} \{ \log F(t_i; \theta) \} \) and show that (usually) the derivative will not equal zero at \( \theta = \theta_0 \).

First we need the density of \( Y_i \). Since \( Y_i = \min \{ T_i, C_i \} \) we see that

\[ P[\min \{ T_i, C_i \} \leq y] = 1 - P[\min \{ T_i, C_i \} > y] \]

\[ = 1 - F(y; \theta_0) G(y) \]

Thus the density of \( Y_i \) is

\[ f(y; \theta_0) G(y) + F(y; \theta_0) g(y). \]

Using the above we have
\begin{equation}
E\left[ \log F(y; \theta) \right] = \int [\log F(y; \theta)] [F(y; \theta_0) F(y) + F(y; \theta_0) g(y)] dy
\end{equation}

To check whether the above is maximum at \( \theta_0 \) we diff the above w.r.t \( \theta \) to give

\[
\frac{\partial}{\partial \theta} E\left[ \log F(y; \theta) \right] = \int \frac{1}{F(y; \theta)} \frac{dF(y; \theta)}{d\theta} \left[ F(y; \theta_0) g(y) + F(y; \theta_0) g(y) \right] dy
\]

At the true parameter \( \theta \) this gives

\[
= \int \frac{dF(y; \theta)}{d\theta} g(y) dy + \int \frac{dF(y; \theta)}{d\theta} F(y; \theta_0) dy.
\]

This term will not = 0 (in general).

\[
= 0 \quad \text{since} \quad \int g(y) dy = 1
\]

Actually \( \neq 0 \)! Yours truly got it wrong... we all make mistakes 😅

Thus we have that

\[
\frac{d}{d\theta} E_{\theta} [L_T(\theta)] = \pi \frac{d}{d\theta} E_{\theta} [\log f(y; \theta)]_{\theta=\theta_0} + (1-\pi) \frac{d}{d\theta} E_{\theta} [\log F(y; \theta)]_{\theta=\theta_0} = 0
\]

\[
\text{not necessarily} = 0.
\]

Since the derivative of expectation of the criterion is not zero at the true parameter \( \theta_0 \), the estimator \( \hat{\theta}_T \) is unlikely to consistently estimate \( \theta_0 \). It will be a biased (in probability) estimator.
Give my huge mistake, shirky that

$$\frac{d}{d\theta} \int \pi(y; \theta) g(y) \, dy = 0 \quad [\text{not quite sure what came over me}].$$

The derivation of the likelihood

$$E \left[ -\frac{d^2 L}{d\theta^2} \right] = E \left[ \left( \frac{d L}{d\theta} \right)^2 \right]$$

so under the assumption.

You should try and see where this derivation falls apart, when \( (\ast) \) is not true!

So I made a huge mistake but it is good practice for you to understand where things work and do not work.
(ii) Now we consider the special case that
\[ g(x) = f(x; \theta_0). \]
In this case the density \( g \)
\[ \min(T, C) \] is
\[ 2 f(x; \theta_0) F(x; \theta_0). \]
Now by substituting this into the calculations in (i) it can be seen that
\[ \frac{\partial}{\partial \theta} \log f(y_i; \theta) \bigg|_{\theta = \theta_0} = 0. \]
Thus in this special case the estimator will be consistent.

(iii) Take second derivatives of
\[ \Pi \equiv \int \log f(t, \theta) \] + \( (1-\Pi) \int \log f(y_i; \theta) \]

give,
\[ \Pi \equiv \int \left\{ -\frac{1}{f(t, \theta)} \left( \frac{\partial f}{\partial \theta} \right)^2 + \frac{1}{f(t, \theta)} \frac{\partial^2 f}{\partial \theta^2} \right\} \]
\[ + (1-\Pi) \int \left\{ -\frac{1}{f(y_i, \theta)} \left( \frac{\partial f}{\partial \theta} \right)^2 + \frac{1}{f(y_i, \theta)} \frac{\partial^2 f}{\partial \theta^2} \right\} \]

Under the assumptions given in (iii)\n\[ \Pi \equiv \int \left\{ -\frac{1}{f(y_i, \theta)} \left( \frac{\partial f}{\partial \theta} \right)^2 \right\} = 0. \]
Thus, \( \theta(x) = \mathcal{F}(x; \theta_0) \) we have

\[
\begin{align*}
\frac{\partial}{\partial \theta} \left[ \frac{1}{\mathcal{F}(T; \theta)} \right] &+ (1-\pi) \mathbb{E} \left[ \frac{1}{\mathcal{F}(Y_i; \theta)} \right] \\
= \mathbb{E} \left[ \frac{1}{\mathcal{F}(T; \theta)} \left( \frac{\partial \mathcal{F}(T; \theta)}{\partial \theta} \right)^2 \right] + (1-\pi) \mathbb{E} \left[ \frac{1}{\mathcal{F}(Y_i; \theta)} \left( \frac{\partial \mathcal{F}(Y_i; \theta)}{\partial \theta} \right)^2 \right] \\
= - \mathbb{E} \left[ \frac{\partial}{\partial \theta} \left( \frac{\mathcal{F}(T; \theta) \log \mathcal{F}(T; \theta) + (1-\pi) \log \mathcal{F}(Y_i; \theta)}{\mathcal{F}(T; \theta)} \right) \right]^2.
\end{align*}
\]

Then, the limiting distribution of the estimator \( \hat{\theta} \) (based on the survival likelihood) is

\[\sqrt{T} (\hat{\theta} - \theta_0) \rightarrow N \left( 0, I_1 \right),\]

where

\[I_1 = \mathbb{E} \left[ \frac{1}{\mathcal{F}(T; \theta)} \left( \frac{\partial \mathcal{F}(T; \theta)}{\partial \theta} \right)^2 \right] + (1-\pi) \mathbb{E} \left[ \frac{1}{\mathcal{F}(Y_i; \theta)} \left( \frac{\partial \mathcal{F}(Y_i; \theta)}{\partial \theta} \right)^2 \right].\]

(iv) An alternative method is to maximise the non-censored part of the density, which is

\[L_2(\theta) = \sum_i s_i \log \mathcal{F}(T_i; \theta).\]

The limiting distribution of the estimator \( \hat{\theta}_2 \) is

\[\sqrt{T} (\hat{\theta}_2 - \theta_0) \rightarrow N \left( 0, I_2 \right),\]

where
\[ I_2 = \prod \theta \left\{ \frac{1}{\hat{F}(T; \hat{\theta})} \left( \frac{\partial \hat{\theta}}{\partial \theta} \right)^2 \right\}. \]

We can easily see that \( I_1 > I_2 \), thus the limiting variance of \( \hat{\theta}_1 \) is smaller than \( \hat{\theta}_2 \). Hence in the special case that \( G(x) = \hat{F}(x; \theta) \), it makes sense to use the survival likelihood to estimate \( \theta_0 \).