Universidad de Talca, FEN

Handout: Estimation of Linear Regression Model by ML

Professor: Mauricio Sarrias

2020

1 Consistency of Maximum Likelihood Estimator

Let the model be $y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta}_0 + \epsilon_i$, such that $\epsilon_i | \boldsymbol{x}_i \sim \mathrm{N}(0, \sigma_0^2)$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\intercal, \sigma^2)^\intercal$ and $\boldsymbol{w}_i = (y_i, \boldsymbol{x}_i^\intercal)^\intercal$, then the log-likelihood function is given by:

$$\begin{aligned}\log L(\boldsymbol{\theta}; \boldsymbol{w}) &= \sum_{i=1}^n \left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\frac{\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}\right)^2}{\sigma^2}\right] \\ &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\frac{\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\intercal\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)}{\sigma^2}\end{aligned} \tag{1.1}$$

The parameter space $\boldsymbol{\Theta}$ is $\mathbb{R}^K \times \mathbb{R}_{++}$, where K is the dimension of $\boldsymbol{\beta}$ and $\mathbb{R}_{++}$ is the set of positive real numbers reflecting the a priori restriction that $\sigma_0^2 > 0$.

We will show that the ML estimators given by:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\intercal\boldsymbol{X}\right)\boldsymbol{X}^\intercal\boldsymbol{y}$$

and

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^\intercal\widehat{\boldsymbol{\varepsilon}}}{n}$$

where $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ are consistent using our *Consistency of Maxima with Compact Parameter Space* Theorem (See slides).

1. **Compact parameter space**: Define a compact parameter space $\boldsymbol{\Theta}$ by:

$$c_1 \leq \sigma_2^2 \leq c_2, \quad \boldsymbol{\beta}^\intercal \boldsymbol{\beta} \leq c_3$$

where c_1 is a small positive constant and c_2 and c_3 are large positive constants, and assume that $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\intercal, \sigma_0^2)^\intercal$ is an interior point of $\boldsymbol{\Theta} = \mathbb{R}^K \times \mathbb{R}_{++}$.

2. **Uniform convergence**: Let $Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n L(\boldsymbol{\theta}; y_i|\boldsymbol{x}_i)$, where

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n L(\boldsymbol{\theta}; y_i|\boldsymbol{x}_i) &= \frac{1}{n}\sum_{i=1}^n \left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\frac{\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}\right)^2}{\sigma^2}\right] \\ &= -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\frac{1}{n}\sum_{i=1}^n\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}\right)^2\end{aligned} \tag{1.2}$$

We need to show that the assumptions of ULLN theorem (see slides) are met. Continuity and iid assumptions are met. We'll mainly prove that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n L(\boldsymbol{\theta}; y_i|\boldsymbol{x}_i)\right]$ exists. Taking expectation over Equation (1.2), we obtain:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n L(\boldsymbol{\theta}; y_i|\boldsymbol{x}_i)\right] &= \mathbb{E}\left[-\frac{1}{2}\log(2\pi\sigma^2)\right] - \frac{1}{2\sigma^2}\mathbb{E}\left[\frac{1}{n}\sum\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}\right)^2\right] \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n(\boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0 + \epsilon_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta})^2\right] \quad \because y_i = \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0 + \epsilon_i \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n\left(\boldsymbol{x}_i^\intercal(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \epsilon_i\right)^2\right] \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n\left((\boldsymbol{\beta}_0 - \boldsymbol{\beta})\boldsymbol{x}_i\boldsymbol{x}_i^\intercal(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + 2\boldsymbol{x}_i^\intercal(\boldsymbol{\beta}_0 - \boldsymbol{\beta})\epsilon_i + \epsilon_i^2\right)\right] \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{n}\sum_{i=1}^n\left[(\boldsymbol{\beta}_0 - \boldsymbol{\beta})^\intercal\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \sigma_0^2\right] \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{n}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})^\intercal\sum_{i=1}^n\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) - \frac{1}{2}\frac{\sigma_0^2}{\sigma^2}\end{aligned}$$

Thus, the expectation exists if and only if $\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)$ exits. Using the ULLN we state that:

$$\frac{1}{n}\sum_{i=1}^n L(\boldsymbol{\theta}; y_i|\boldsymbol{x}_i) \xrightarrow{p} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})^\intercal\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) - \frac{1}{2}\frac{\sigma_0^2}{\sigma^2} = \mathbb{E}\left[L(\boldsymbol{\theta}; y_i|\boldsymbol{x}_i)\right]$$

uniformly for some fixed parameter value $\boldsymbol{\theta}$, since $\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\intercal \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)$ by assuming that $\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)$ exists.

3. **Identification**: We need to show that $\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta}) \neq \log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta}_0)$ with positive probability if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Recall that:

$$\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\frac{\left(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}\right)^2}{\sigma^2}\right].$$

Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\intercal, \sigma_0^2)^\intercal$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\intercal, \sigma^2)^\intercal$. It can be observed that $\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta}) \neq \log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta}_0)$ with positive probability if $\sigma^2 \neq \sigma_0^2$. Now consider the case where $\sigma^2 = \sigma_0^2$, but $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. If this is the case, then:

$$\mathbb{E}\left[\left(\boldsymbol{x}_i^\intercal\boldsymbol{\beta} - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0\right)^2\right] = (\boldsymbol{\beta}_0 - \boldsymbol{\beta})^\intercal\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) > 0$$

This implies that $\boldsymbol{x}_i^\intercal\boldsymbol{\beta} \neq \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0$ with positive probability. Note if $\boldsymbol{x}_i^\intercal\boldsymbol{\beta} = \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0$ with probability 1, then $\mathbb{E}\left[\left(\boldsymbol{x}_i^\intercal\boldsymbol{\beta} - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0\right)^2\right]$ will be zero. Thus $y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta} \neq y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0$ with positive probability, implying that the identification condition is satisfied.

Another way is the following. Note that the conditional expectation of the conditional log-likelihood function is:

$$\begin{aligned}
\mathbb{E}\left[\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta})|\boldsymbol{x}_i\right] &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{\mathbb{E}\left[(y_i - \boldsymbol{x}_i^\intercal\boldsymbol{\beta})^2\right]}{2\sigma^2} \\
&= -\frac{1}{2}\left[\log(2\pi\sigma^2) + \frac{\sigma_0^2 + (\boldsymbol{x}_i^\intercal\boldsymbol{\beta} - \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0)^2}{\sigma^2}\right]
\end{aligned}$$

where we use the fact that $\mathbb{E}(X-\mu)^2 = \mathrm{Var}(X) + (\mathbb{E}(X) - \mu)^2$. Note that if $\boldsymbol{x}_i'\boldsymbol{\beta} = \boldsymbol{x}_i'\boldsymbol{\beta}_0$ and $\sigma^2 = \sigma_0^2$. Then:

$$\mathbb{E}\left[\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta})|\boldsymbol{x}_i\right] = -\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}$$

So, it must be the case that $\mathbb{E}\left[\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta})|\boldsymbol{x}_i\right]$ is uniquely maximized at $\boldsymbol{x}_i^\intercal\boldsymbol{\beta} = \boldsymbol{x}_i^\intercal\boldsymbol{\beta}_0$ and $\sigma^2 = \sigma_0^2$ ($\mathbb{E}\left[\log f(y_i|\boldsymbol{x}_i;\boldsymbol{\theta})|\boldsymbol{x}_i\right]$ does not depend on the parameters) if $\boldsymbol{X}$ is full-column rank.

Thus, we can say that $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\intercal, \widehat{\sigma}^2)^\intercal \xrightarrow{p} \boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\intercal, \sigma_0^2)^\intercal$.

2 Score function and Hessian Function

Now we derive the score and Hessian function for the Linear Regression. We also will check some of the properties.

Recall that the elements of the score function are:

$$
\begin{aligned}
\frac{\partial \log L_i(\boldsymbol{\theta})}{\partial \widehat{\boldsymbol{\beta}}} &= -\frac{2\left(y_i - \boldsymbol{x}_i^\intercal \widehat{\boldsymbol{\beta}}\right)}{2\widehat{\sigma}^2}(-\boldsymbol{x}_i) \\
&= -\frac{\left(y_i - \boldsymbol{x}_i^\intercal \widehat{\boldsymbol{\beta}}\right)\boldsymbol{x}_i}{\widehat{\sigma}^2} \\
&= -\frac{\boldsymbol{x}_i \widehat{\epsilon}_i}{\widehat{\sigma}^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log L_i(\boldsymbol{\theta})}{\partial \widehat{\sigma}^2} &= -\frac{1}{2}\frac{1}{2\pi\widehat{\sigma}^2}2\pi - \frac{\left(y_i - \boldsymbol{x}_i^\intercal \widehat{\boldsymbol{\beta}}\right)^2}{2}\left(-\widehat{\sigma}^{-4}\right) \\
&= -\frac{1}{2\widehat{\sigma}^2} + \frac{\widehat{\epsilon}_i^2}{2\widehat{\sigma}^4}
\end{aligned}
$$

So that the score function for individual i is:

$$
\boldsymbol{s}(\boldsymbol{w}_i; \widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{\boldsymbol{x}_i \widehat{\epsilon}_i}{\widehat{\sigma}^2} \\ -\frac{1}{2\widehat{\sigma}^2} + \frac{\widehat{\epsilon}_i^2}{2\widehat{\sigma}^4} \end{pmatrix} \tag{2.1}
$$

where $\boldsymbol{w}_i = (y_i, \boldsymbol{x}_i^\intercal)^\intercal$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\intercal, \sigma^2)^\intercal$ and $\widehat{\epsilon}_i \equiv y_i - \boldsymbol{x}_i^\intercal \boldsymbol{\beta}$. The score function for the whole sample is:

$$
\begin{aligned}
\boldsymbol{s}(\boldsymbol{w}; \widehat{\boldsymbol{\theta}}) &= \begin{pmatrix} \sum_{i=1}^n \frac{\boldsymbol{x}_i \widehat{\epsilon}_i}{\widehat{\sigma}^2} \\ -\sum_{i=1}^n \frac{1}{2\widehat{\sigma}^2} + \sum_{i=1}^n \frac{\widehat{\epsilon}_i^2}{2\widehat{\sigma}^4} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\widehat{\sigma}^2}\sum_{i=1}^n \boldsymbol{x}_i \widehat{\epsilon}_i \\ -\frac{n}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}\sum_{i=1}^n \widehat{\epsilon}_i^2 \end{pmatrix}
\end{aligned} \tag{2.2}
$$

An important property is the *Score Identity*, which states that $\mathbb{E}(\boldsymbol{s}(\boldsymbol{w}; \boldsymbol{\theta}_0)) = \mathbf{0}$, which can be proven as follows:

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta}_0)|\boldsymbol{x}_i) &= \begin{pmatrix} \mathbb{E}\left(\mathbb{E}\left[\frac{\boldsymbol{x}_i \epsilon_i}{\sigma_0^2}|\boldsymbol{x}_i\right]\right) \\ \mathbb{E}\left(\mathbb{E}\left[-\frac{1}{2\sigma_0^2} + \frac{\epsilon_i^2}{2\sigma_0^4}|\boldsymbol{x}_i\right]\right) \end{pmatrix} \\
&= \begin{pmatrix} \mathbb{E}\left(\frac{1}{\sigma_0^2}\boldsymbol{x}_i \mathbb{E}\left[\epsilon_i|\boldsymbol{x}_i\right]\right) \\ \mathbb{E}\left(-\frac{1}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\mathbb{E}\left[\epsilon_i^2|\boldsymbol{x}_i\right]\right) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0} \\ \mathbb{E}\left(-\frac{1}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\sigma_0^2\right) \end{pmatrix} \\
&= \mathbf{0}
\end{aligned} \tag{2.3}
$$

So for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ the $\widehat{\epsilon}_i$ in these expressions can be replaced by ϵ_i. Recall also that in the linear regression model, $\mathbb{E}(\epsilon_i|\boldsymbol{x}_i) = 0$.

You should be able to obtain that:

$$\boldsymbol{s}(\boldsymbol{w}_i;\boldsymbol{\theta})^\intercal \boldsymbol{s}(\boldsymbol{w}_i;\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\sigma^4}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal \widehat{\epsilon}_i^2 & -\frac{1}{2\sigma^4}\boldsymbol{x}_i \cdot \widehat{\epsilon}_i + \frac{1}{2\sigma^6}\boldsymbol{x}_i \cdot \widehat{\epsilon}_i^3 \\ -\frac{1}{2\sigma^4}\boldsymbol{x}_i^\intercal \cdot \widehat{\epsilon}_i + \frac{1}{2\sigma^6}\boldsymbol{x}_i^\intercal \cdot \widehat{\epsilon}_i^3 & \frac{1}{4\sigma^4} - \frac{1}{2\sigma^6}\widehat{\epsilon}_i^2 + \frac{1}{4\sigma^8}\widehat{\epsilon}_i^4 \end{pmatrix} \tag{2.4}$$

Similarly, it is easy to show that:

$$\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}) = \begin{pmatrix} -\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\intercal}{\sigma^2} & -\frac{1}{\sigma^4}\boldsymbol{x}_i \cdot \widehat{\epsilon}_i \\ -\frac{1}{\sigma^4}\boldsymbol{x}_i^\intercal \cdot \widehat{\epsilon}_i & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\widehat{\epsilon}_i^2 \end{pmatrix} \tag{2.5}$$

3 Asymptotic Normality

We will check the assumptions of *Asymptotic normality of Conditional ML* Theorem (See slides).

1. Assume that $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta} = \mathbb{R}^K \times \mathbb{R}_{++}$.
2. We have shown that $f(\boldsymbol{w}_i : \boldsymbol{\theta}_0)$ is twice continuously differentiable in $\boldsymbol{\theta}$.
3. We have shown that $\mathbb{E}(\boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})) = \mathbf{0}$ in Equation (2.3). Now, we prove that $-\mathbb{E}\left[\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\right] = \mathbb{E}\left[\boldsymbol{s}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)^\intercal\right]$.

   Taking the expectation over Equation 2.5, we obtain:

$$\begin{aligned} -\mathbb{E}\left[\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)|\right] &= -1 \cdot \begin{pmatrix} -\mathbb{E}\left[\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\intercal}{\sigma_0^2}\right] & -\mathbb{E}\left[\frac{1}{\sigma_0^4}\boldsymbol{x}_i \cdot \epsilon_i\right] \\ -\mathbb{E}\left[\frac{1}{\sigma_0^4}\boldsymbol{x}_i^\intercal \cdot \epsilon_i\right] & \mathbb{E}\left[\frac{1}{2\sigma_0^4} - \frac{1}{\sigma_0^6}\epsilon_i^2\right] \end{pmatrix} \\ &= -1 \cdot \begin{pmatrix} -\frac{1}{\sigma_0^2}\mathbb{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right] & -\frac{1}{\sigma_0^4}\mathbb{E}\left[\boldsymbol{x}_i \cdot \epsilon_i\right] \\ -\frac{1}{\sigma_0^4}\mathbb{E}\left[\boldsymbol{x}_i^\intercal \cdot \epsilon_i\right] & \frac{1}{2\sigma_0^4} - \frac{1}{\sigma_0^6}\mathbb{E}\left[\epsilon_i^2\right] \end{pmatrix} \\ &= -1 \cdot \begin{pmatrix} -\frac{1}{\sigma_0^2}\mathbb{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right] & \mathbf{0} \\ \mathbf{0}^\intercal & \frac{1}{2\sigma_0^4} - \frac{1}{\sigma_0^6}\sigma_0^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbb{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right] & \mathbf{0} \\ \mathbf{0}^\intercal & \frac{1}{2\sigma_0^4} \end{pmatrix} \end{aligned} \tag{3.1}$$

   Also, since $\epsilon_i \sim \mathrm{N}(0, \sigma_0^2)$, we have $\mathbb{E}(\epsilon_i^3) = 0$ and $\mathbb{E}(\epsilon_i^4) = 3\sigma_0^4$. Using these relations, you should able to prove that:

$$-\mathbb{E}\left[\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\right] = \mathbb{E}\left[\boldsymbol{s}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)^\intercal\right] = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal) & \mathbf{0} \\ \mathbf{0}^\intercal & \frac{1}{2\sigma_0^4} \end{pmatrix}$$

   Now, we need to show that $\frac{1}{n}\sum_{i=1}^n \boldsymbol{H}(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbb{E}\left[\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\right]$. Since $\widehat{\boldsymbol{\theta}}$ is consistent, note that:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\intercal}{\widehat{\sigma}^2} \xrightarrow{p} \frac{1}{\sigma_0^2}\mathbb{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right)$$

Since $\widehat{\epsilon}_i = \epsilon_i - \boldsymbol{x}_i^\intercal\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$, then

$$\begin{aligned}\frac{1}{\widehat{\sigma}^4}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\widehat{\epsilon}_i &= \frac{1}{\widehat{\sigma}^4}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\left(\epsilon_i - \boldsymbol{x}_i^\intercal\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right)\\ &\xrightarrow{p} \frac{1}{\sigma_0^4}\mathbb{E}\left(\boldsymbol{x}_i\epsilon_i\right) - \frac{1}{\sigma_0^4}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\mathbb{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right)\\ &\xrightarrow{p} \mathbf{0}\end{aligned}$$

assuming that $\mathbb{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right)$ exists and $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Finally:

$$\begin{aligned}\frac{1}{2\widehat{\sigma}^4} - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\sigma}^6}\widehat{\epsilon}_i^2 &= \frac{n}{2\widehat{\sigma}^4} - \frac{1}{\widehat{\sigma}^6}\frac{1}{n}\sum_{i=1}^{n}\left(\epsilon_i - \boldsymbol{x}_i^\intercal\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right)^2\\ &= \frac{1}{2\widehat{\sigma}^4} - \frac{1}{\widehat{\sigma}^6}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 + \frac{1}{\widehat{\sigma}^6}\frac{1}{n}\sum_{i=1}^{n}2\boldsymbol{x}_i^\intercal\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) - \frac{1}{\widehat{\sigma}^6}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)^\intercal\boldsymbol{x}_i\boldsymbol{x}_i^\intercal\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\\ &\xrightarrow{p} \frac{1}{2\sigma_0^4} - \frac{1}{\sigma_0^6}\mathbb{E}(\epsilon_i^2)\\ &\xrightarrow{p} \frac{1}{2\sigma_0^4} - \frac{1}{\sigma_0^6}\sigma_0^2\\ &\xrightarrow{p} -\frac{1}{2\sigma_0^4}\end{aligned}$$

4. $\mathbb{E}\left[\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\right]$ is nonsingular. Using Equation 3.1, if $\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal)$ is nonsingular, then $\mathbb{E}\left[\boldsymbol{H}(\boldsymbol{w}_i;\boldsymbol{\theta}_0)\right]$ is nonsingular.

Thus, we can say that:

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathrm{N}\left[\mathbf{0}, \begin{pmatrix}\frac{1}{\sigma_0^2}\mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal) & \mathbf{0}\\ \mathbf{0}' & \frac{1}{2\sigma_0^4}\end{pmatrix}^{-1}\right]$$