The result (5.44) essentially proves the Gauss-Markov Theorem, since it implies that

$$\begin{aligned}
&E(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \\
&= E\Big(\big((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u} + \boldsymbol{C}\boldsymbol{u}\big)\big((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u} + \boldsymbol{C}\boldsymbol{u}\big)^\top\Big) \\
&= \sigma_0^2\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1} + \sigma_0^2\boldsymbol{C}\boldsymbol{C}^\top.
\end{aligned} \tag{5.45}$$

Thus the difference between the covariance matrices of $\check{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ is $\sigma_0^2\boldsymbol{C}\boldsymbol{C}^\top$, which is a positive semidefinite matrix. Notice that the assumption that $E(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma_0^2\mathbf{I}$ is crucial here. If instead we had $E(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Omega}$, with $\boldsymbol{\Omega}$ an arbitrary $n \times n$ positive definite matrix, the last line of (5.45) would be

$$\begin{aligned}
&\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1} \\
&\quad + \boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}^\top + \big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{C}^\top + \boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1},
\end{aligned}$$

and we could draw no conclusion about the relative efficiency of $\hat{\boldsymbol{\beta}}$ and $\check{\boldsymbol{\beta}}$.

As a simple example of the Gauss-Markov Theorem in action, suppose that $\check{\boldsymbol{\beta}}$ is the OLS estimator obtained by regressing $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\boldsymbol{Z}$ jointly, where $\boldsymbol{Z}$ is a matrix of regressors such that $E(\boldsymbol{y}\,|\,\boldsymbol{X}, \boldsymbol{Z}) = E(\boldsymbol{y}\,|\,\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}$. Since the information that $\boldsymbol{Z}$ does not belong in the regression is being ignored when we construct $\check{\boldsymbol{\beta}}$, the latter must in general be inefficient. Using the FWL Theorem, we find that

$$\check{\boldsymbol{\beta}} = \big(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{y}, \tag{5.46}$$

where, as usual, $\boldsymbol{M}_Z$ is the matrix that projects orthogonally onto $\mathcal{S}^\perp(\boldsymbol{Z})$. If we write $\check{\boldsymbol{\beta}}$ as in (5.42), we obtain

$$\begin{aligned}
\check{\boldsymbol{\beta}} &= \big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{y} + \big((\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{M}_Z - (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\big)\boldsymbol{y} \\
&= \big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{y} + \big(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}\big)^{-1}\big(\boldsymbol{X}^\top\boldsymbol{M}_Z - \boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\big)\boldsymbol{y} \\
&= \big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{y} + \big(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}\big)^{-1}\Big(\boldsymbol{X}^\top\boldsymbol{M}_Z\big(\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\big)\Big)\boldsymbol{y} \\
&= \big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{y} + \big(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{M}_X\boldsymbol{y} \\
&= \hat{\boldsymbol{\beta}} + \boldsymbol{C}\boldsymbol{y}.
\end{aligned} \tag{5.47}$$

Thus, in this case, the matrix $\boldsymbol{C}$ is the matrix $(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{M}_X$. We see that the inefficient estimator $\check{\boldsymbol{\beta}}$ is equal to the efficient estimator $\hat{\boldsymbol{\beta}}$ plus a random component which is uncorrelated with it. That $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{C}\boldsymbol{y}$ are uncorrelated follows from the fact (required for $\boldsymbol{C}\boldsymbol{y}$ to have mean zero) that $\boldsymbol{C}\boldsymbol{X} = \mathbf{0}$, which is true because $\boldsymbol{M}_X$ annihilates $\boldsymbol{X}$. Further, we see that

$$\begin{aligned}
E(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top &= \sigma_0^2\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1} \\
&\quad + \sigma_0^2\big(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{M}_X\boldsymbol{M}_Z\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{M}_Z\boldsymbol{X}\big)^{-1}.
\end{aligned} \tag{5.48}$$