5. As we saw, for model $\mathcal{M}_\gamma$, the corresponding backprojection prior for full model $\mathcal{M}$ is:

$$\tilde{\beta}_\gamma = (X_\gamma^T X_\gamma)^{-1} X_\gamma^T X \tilde{\beta} \tag{1}$$

with $\tilde{\beta}$ is the prior mean of β in $\mathcal{M}$. The prior distribution for $\tilde{\beta}_\gamma$ in the first level G-prior setting should not be simply partition in a $(X_\gamma, X_{-\gamma})$ way. Instead, using the projection matrix $P = (X_\gamma^T X_\gamma)^{-1} X_\gamma^T X$, we have:

$$\beta_\gamma | \gamma, \sigma^2 \sim \mathcal{N}_{q_\gamma+1}(\tilde{\beta}_\gamma, \tilde{\Sigma}) \tag{2}$$

where

$$\tilde{\Sigma} = Pc\sigma^2(X^T X)^{-1} P^T = c\sigma^2 (X_\gamma^T X_\gamma)^{-1} X_\gamma^T [X(X^T X)^{-1} X^T] X_\gamma (X_\gamma^T X_\gamma)^{-1}. \tag{3}$$

This is different from the prior variance $c\sigma^2(X_\gamma^T X_\gamma)^{-1}$. This is equivalent to shrink your original variance hyperparameter by a fact of hat matrix $H = X(X^T X)^{-1} X^T$.

By orthogonal design assumption: $X_\gamma^T X_{-\gamma} = X_{-\gamma}^T X_\gamma = \mathbf{0}$, and expanding $\tilde{\Sigma}$, we can easily derive:

$$\tilde{\Sigma} = c\sigma^2 (X_\gamma^T X_\gamma)^{-1} (H_\gamma + H_{-\gamma}) \tag{4}$$

where $H_\gamma = X_\gamma (X_\gamma^T X_\gamma)^{-1} X_\gamma^T$ and $H_{-\gamma} = X_{-\gamma} (X_{-\gamma}^T X_{-\gamma})^{-1} X_{-\gamma}^T$, i.e. hat matrix of corresponding blocks.

An important notice is that if assuming $X = (X_\gamma, X_{-\gamma})$ is an orthogonal design then we have

$$\begin{aligned} \tilde{\Sigma} &= c\sigma^2 (X_\gamma^T X_\gamma)^{-1} (H_\gamma + H_{-\gamma}) \\ &= c\sigma^2 (X_\gamma^T X_\gamma) \end{aligned}$$

which exactly what you got in the book. However, in general case where orthogonal design is not applicable, then dependency between X_γ and $X_{-\gamma}$ may be introduced. In this more general settings, Zellner's G-prior should be exactly as Eq.(1).

Cosequently, the joint prior, marginal density of y and γ are all not incorrect in the rest of Section 3.5.2.

To mitigate the unclearness, we could just assume the orthogonal design without loss of generality to a wider settings. This assumption, however, would be proper for comparing just a nested small model to the full model.

For the model selection issue, e.g. using stochastic search, this orthogonal design would probably harm for further inference! (You have to assume orthogonality for every possible model partition!) Hence, I think, more formally, presenting the general results to the readers is a better idea.