# Joint Specification of Model Space and Parameter Space Prior Distributions

Petros Dellaportas*, Jonathan J. Forster† and Ioannis Ntzoufras*

*Abstract:* We consider the specification of prior distributions for Bayesian model comparison, focusing on regression-type models. We propose a particular joint specification of the prior distribution across models so that sensitivity of posterior model probabilities to the dispersion of prior distributions for the parameters of individual models (Lindley's paradox) is diminished. We illustrate the behavior of inferential and predictive posterior quantities in linear and log-linear regressions under our proposed prior densities with a series of simulated and real data examples.

*Keywords:* Bayesian inference, BIC, Generalised linear models, Lindley's Paradox, Model averaging, Regression models.

## A Electronic Appendix - Detailed Additional Results for the Simulations of Section 6.2

This is a web supplement to the paper "*Joint Specification of Model Space and Parameter Space Prior Distributions*", by Dellaportas, Forster and Ntzoufras (2011) which is available at `http://stat-athens.aueb.gr/~jbn/papers/paper24.htm`. In this electronic Appendix we present detailed results after generating 100 datasets using the sampling schemes (22) and (23) described in Section 6.2.

### A.1 First Simulation Scheme

In this first simulation study, we use the sampling scheme (22) of Section 6.2. Therefore we have generated 100 datasets of $n = 50$ observations of 15 standardized independent normal covariates $X_j,\ j = 1, \dots, 15$, with responses $Y \sim N(\ X_4 + X_5,\ 2.5^2\ )$ .

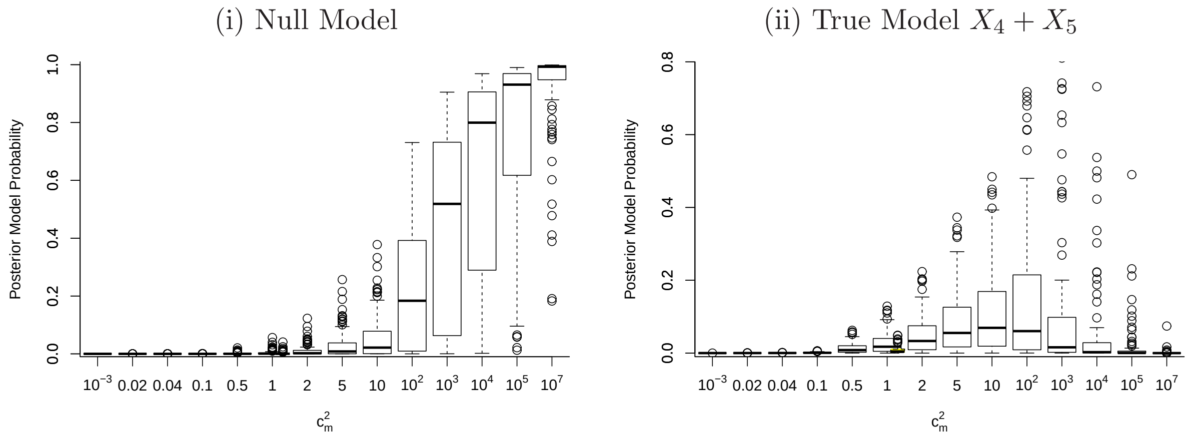
Figures 6 and 7 depict the posterior model and variable inclusion probabilities using boxplots under repeating sampling (i.e. 100 random samples). In Figure 6 the posterior model probabilities for the null and the true model under the prior setups described in Section 6.2 are presented while in Figure 7 the posterior variable inclusion probabilities for two indicative covariates are presented (the picture for all non-important covariates is similar to $X_1$ while $X_5$ has the same behaviour as $X_4$ since their true coefficients are equal). The behaviour of all methods is similar to the one for

*Department of Statistics, Athens University of Economics and Business, 76 Patission Str, 10434 Athens, Greece, emails: petros@aueb.gr, ntzoufras@aueb.gr

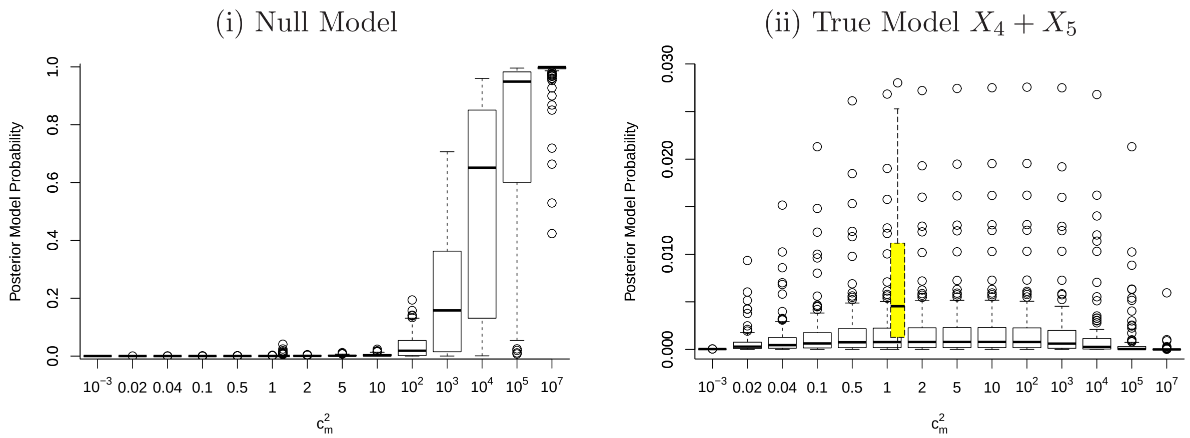
†Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton, SO17 1BJ, UK, email: J.J.Forster@soton.ac.uk

the single case study analysed in Section 6.2 with our approach being robust for all values of $c_m^2$ while the Zellner's g-prior with the discrete prior on model space is highly sensitive to the choice of $c_m^2$. The hyper-g (with the discrete prior on model space) is extremely robust for a wide range of $a$. Nevertheless, Lindley's paradox eventually appears as $a$ approaches 2.

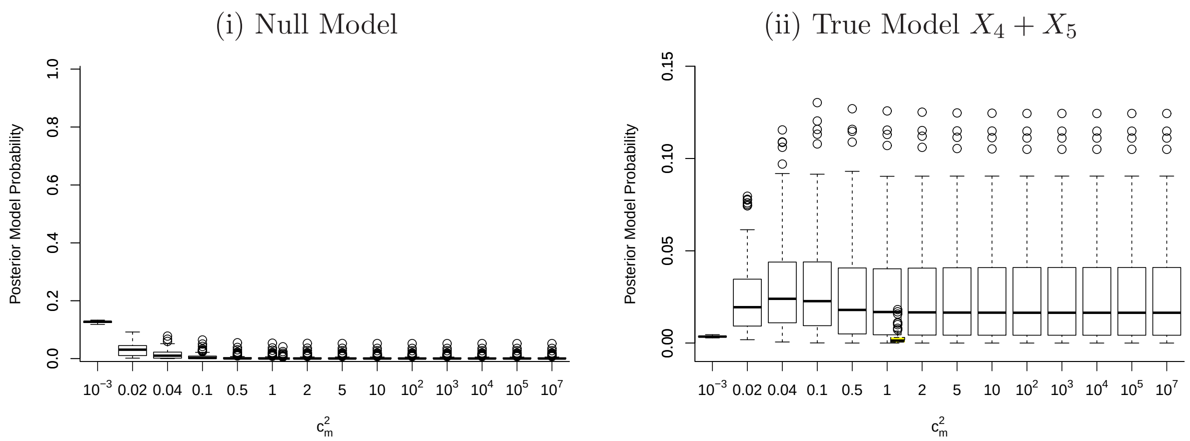
It is notable that, for the hyper-g prior, the true model $(X_4 + X_5)$ has a remarkably low posterior probability ($< 0.005$) for the majority of the generated datasets while the same posterior probabilities are about 5 times higher when using the Zellner's g-prior with the adjusted prior on model space. The same effect is more evident when looking at posterior variable inclusion probabilities where the hyper-g prior gives rather high values to non-important covariates (with median around 0.4) while the corresponding values when using the Zellner's g-prior with the adjusted prior on model space are $\leq 0.2$. The Zellner and Siow prior using discrete prior on model space is also depicted in all Figures with a grey boxplot next to the value for $c_m^2 = 1$. This prior setup results in a posterior distribution with lower model uncertainty and higher posterior probabilities of the true model than the corresponding ones using the hyper-g prior. Nevertheless, these probabilities are still much lower than the ones obtained under the Zellner's g-prior with the adjusted prior on model space.




(a) Zellner's g-prior with uniform prior on model space



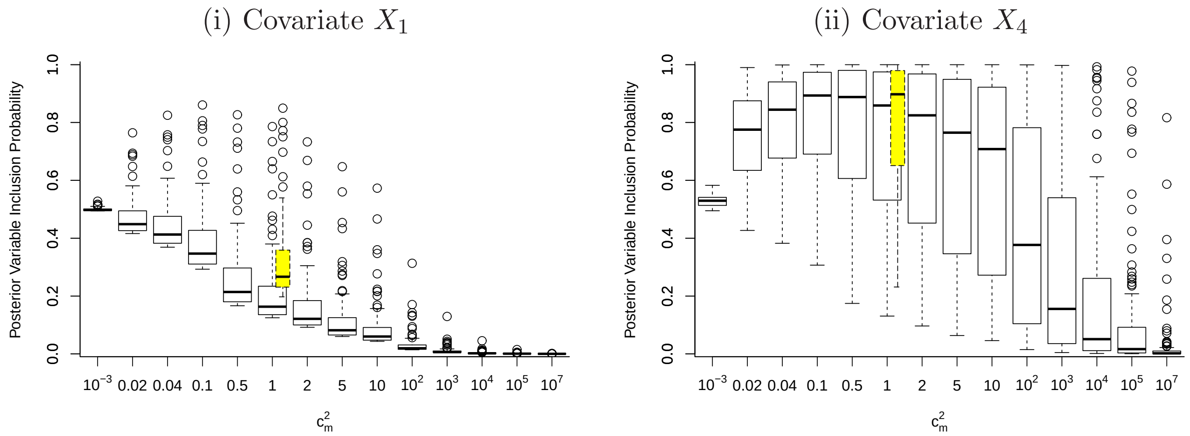

(b) Hyper g-prior with uniform prior on model space



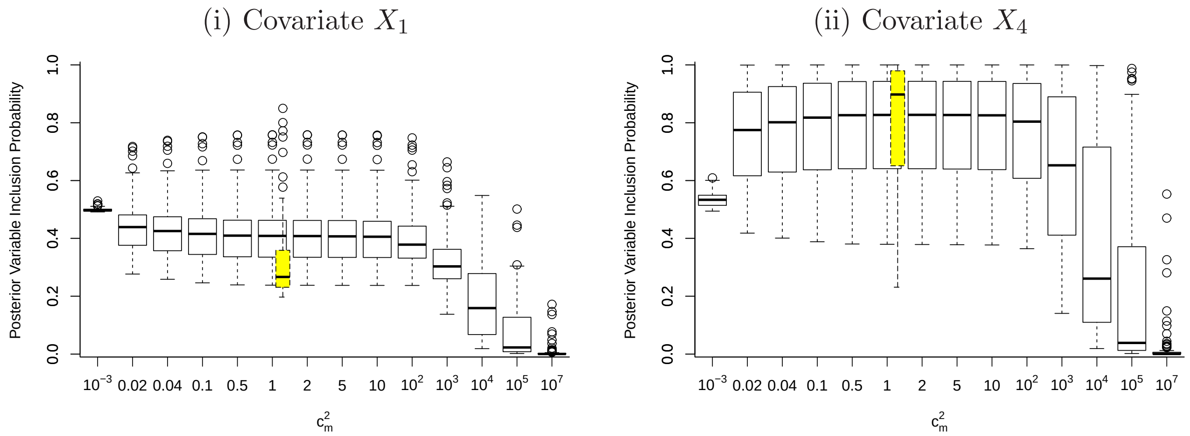

(c) Zellner's g-prior with adjusted prior on model space

Grey (yellow) boxplots (between $c_m^2 = 1$ and $c_m^2 = 2$): Zellner and Siow prior with uniform prior on model space.

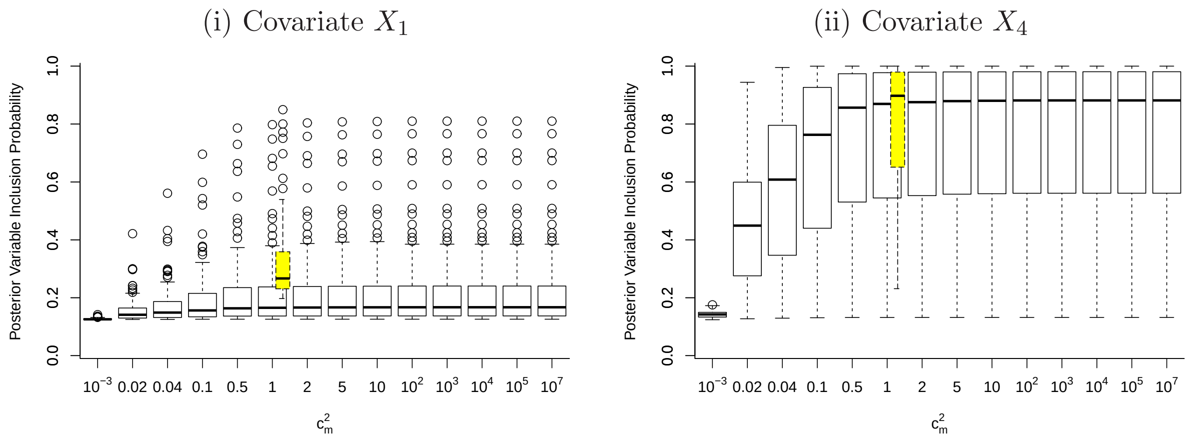
Figure 6: Boxplots of posterior probabilities of the null and the true model under different prior dispersions over 100 simulated datasets for the sampling scheme (22) of Section 6.2.




(a) Zellner's g-prior with discrete prior on model space




(b) Hyper g-prior with discrete prior on model space




(c) Zellner's g-prior with adjusted prior on model space

Grey (yellow) boxplots (between $c_m^2 = 1$ and $c_m^2 = 2$): Zellner and Siow prior using discrete prior on model space.

Figure 7: Boxplots of posterior variable inclusion probabilities under different prior dispersions over 100 simulated datasets for the sampling scheme (22) of Section 6.2.

## A.2 Second Simulation Scheme: Nott and Kohn simulation study

Here illustrate the behaviour of the adjusted prior on model space by considering 100 simulated datasets with the sampling scheme of Nott and Kohn (2005) as described in (23) of Section 6.2. Hence, each data-set consists of $n = 50$ observations and $p = 15$ covariates with the first 10 covariates are generated from a standardized Normal distribution while the rest (i.e. for $j = 11, \ldots, 15$) using expression $X_j \sim N(\,0.3X_1 + 0.5X_2 + 0.7X_3 + 0.9X_4 + 1.1X_5,\ 1\,)$ and the response from $Y \sim N(\,4 + 2X_1 - X_5 + 1.5X_7 + X_{11} + 0.5X_{13},\ 2.5^2\,)$.

Figure 8 presents the distribution, under repeating sampling (i.e. for 100 simulated datasets), of the posterior model probabilities for four models: the null, the true and two sub-models of it which are frequently indicated as the two best ones when using the Zellner's g-prior with $c_m^2 = 1$ and uniform prior on model space. Figure 9 depicts the corresponding distributions of the posterior variable inclusion probabilities for six indicative covariates: for the ones actual non-zero effects and for $X_6$ as representative of the non-important ones. The picture is similar as in the simulation study of Section A.1. Again, the posterior model and variable probabilities are robust for the Zellner's g-prior with the adjusted prior on model space. On the other hand, when adopting the uniform prior on model space then the Zellner's g-prior is highly sensitive to the value of $c_m^2$ while hyper-g is quite robust for a wide range of parameter values. For the latter, the Lindley paradox still makes its appearance for extremely large values of $c_m^2$, i.e. for values of $a$ very close to 2.

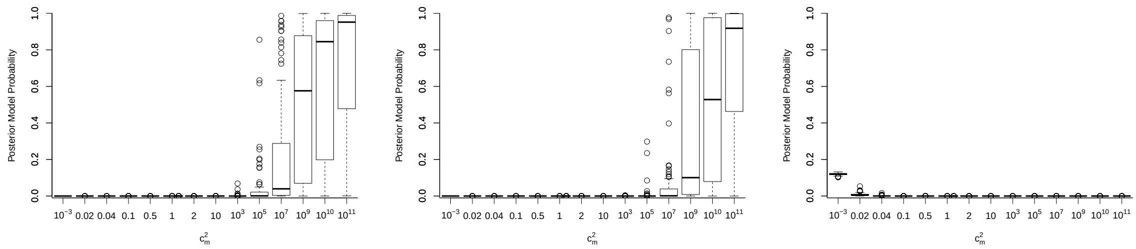
An additional noteworthy difference from the previous simulated exercise, is that, although the hyper-g prior is still supporting the true and the best models with lower posterior probabilities than our approach, the gap between them is considerably smaller possibly due to the inherent model uncertainty of this simulation scheme.

## References for the Appendix

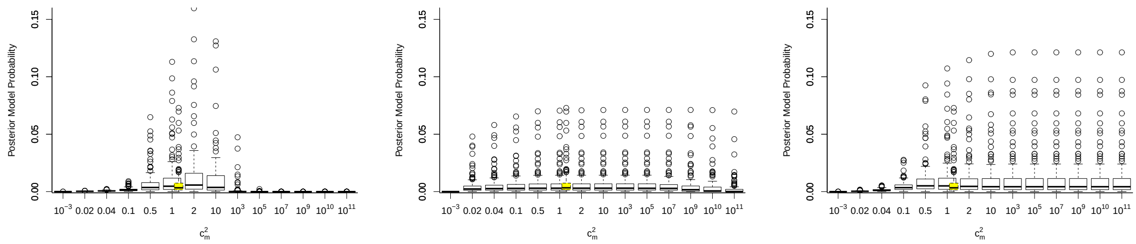

Nott, D. and Kohn, R. (2005). Adaptive sampling for bayesian variable selection, *Biometrika*, **92**, 747–763.

Dellaportas P., Forster, J.J. and Ntzoufras, I. (2011). Joint Specification of Model Space and Parameter Space Prior Distributions, *submitted*, available at `http://stat-athens.aueb.gr/~jbn/papers/paper24.htm`.

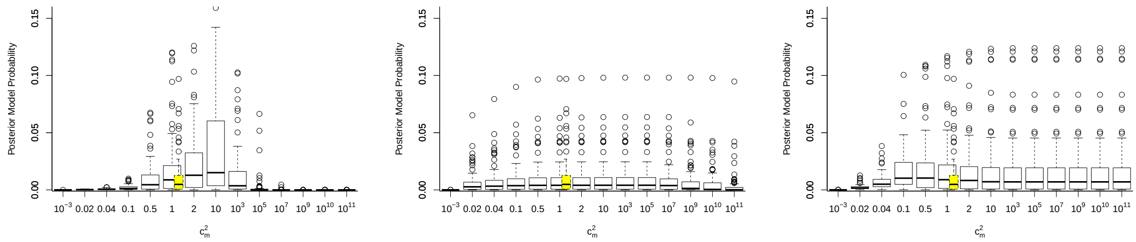
(i) Null Model $m_0$



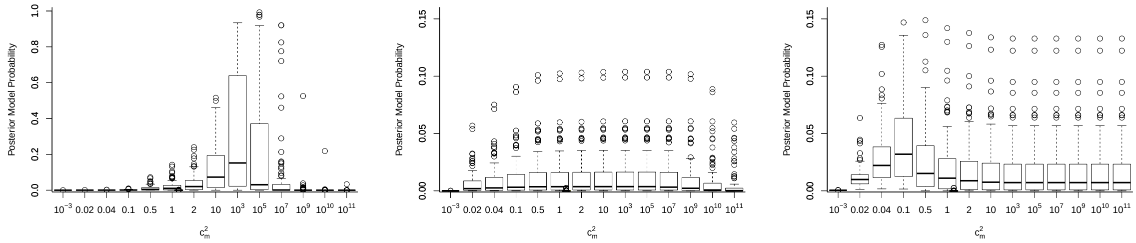

(ii) True Model – $m_1 : X_1 + X_5 + X_7 + X_{11} + X_{13}$




(iii) Model $m_2 : X_1 + X_5 + X_7 + X_{11}$




(iv) Model $m_3 : X_1 + X_7 + X_{11}$

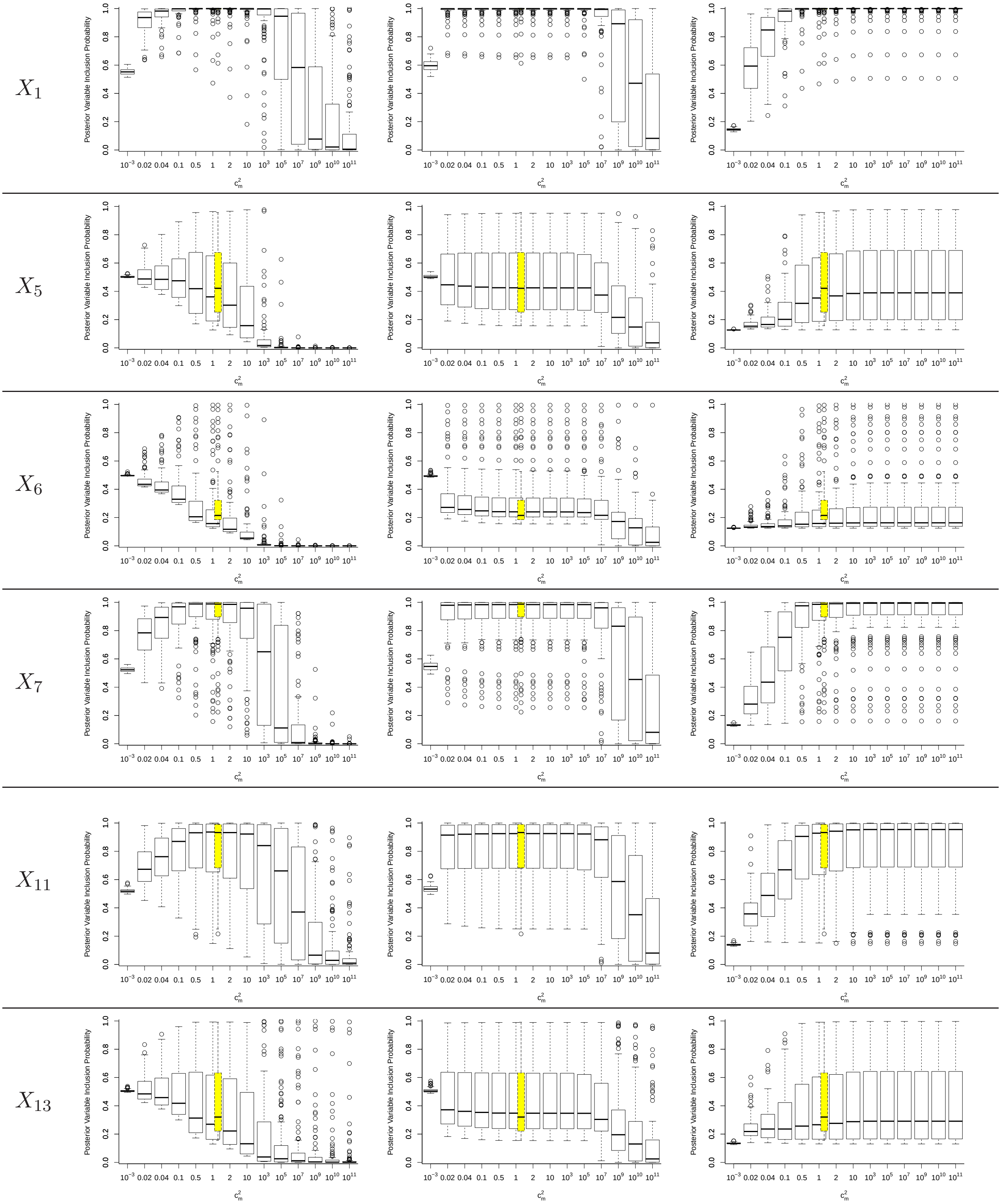



| (a) Zellner's g-prior and uniform on $\mathcal{M}$ | (b) Hyper g-prior and uniform on $\mathcal{M}$ | (c) Zellner's g-prior and adjusted prior on $\mathcal{M}$ |
|---|---|---|

$\mathcal{M}$: model space;

Grey (yellow) boxplots (between $c_m^2 = 1$ and $c_m^2 = 2$): Zellner and Siow prior using uniform on model space.

Figure 8: Boxplots of posterior probabilities of the null, the true model ($m_1 : X_1 + X_5 + X_7 + X_{11} + X_{13}$) and two sub-models of $m_1$ ($m_2 : X_1 + X_5 + X_7 + X_{11}$ and $m_3 : X_1 + X_7 + X_{11}$) under different prior dispersions over 100 simulated datasets for the sampling scheme (23) of Section 6.2.



(a) Zellner's g-prior and uniform on $\mathcal{M}$

(b) Hyper g-prior and uniform on $\mathcal{M}$

(c) Zellner's g-prior and adjusted prior on $\mathcal{M}$

$\mathcal{M}$: model space;

Grey (yellow) boxplots (between $c_m^2 = 1$ and $c_m^2 = 2$): Zellner and Siow prior using uniform prior on model space.

Figure 9: Boxplots of posterior variable inclusion probabilities for selected covariates under different prior dispersions over 100 simulated datasets for the sampling scheme (23) of Section 6.2.