Supplemental material for *Detection of hidden additivity and inference under model uncertainty for unreplicated factorial studies via Bayesian model selection and averaging*

S1 Generalized lasso regularization approach

For a regularization scheme to effectively assess non-additive effects, strict penalty should constrain parameters to reproduce the behavior of the additive model (2). As the penalty is relaxed, non-additive effects enter the model. In this study we allow non-specific interactions of the form in model (1).

Our approach uses a simple custom penalty matrix in the generalized lasso framework (**?**). Following the notation for penalty matrix $\boldsymbol{D}$ in **?** and other notation defined in this supplement's companion paper, we set $\boldsymbol{D} = [\mathbf{0}|\boldsymbol{I}]$, where $\boldsymbol{D}$ is an $n \times n$ matrix, $\mathbf{0}$ is an $n \times (a + b - 1)$ matrix of zeroes, and $\boldsymbol{I}$ is an $n \times (a - 1)(b - 1)$ identity matrix. Choosing this penalty matrix gives the tuning parameter λ the authority to shrink interaction effects toward zero without subjecting the main effects to regularization.

Computation was performed using the *genlasso* package (**??**) in R. In this study, a range of tuning parameters are considered. While out-of-sample cross validation is a common strategy to choose tuning parameters in many applications, that approach is difficult in this setting since omission of even a single point renders some potentially interesting effects non-estimable. Six values of tuning parameter λ were considered. These values of λ are labeled $I = 0.04$, $II = 0.20$, $III = 0.43$, $IV = 0.94$, $V = 2.07$, and $VI = 9.97$ in the companion paper. We selected these values from a larger grid of tuning parameters equally spaced on the log scale during our analysis of simulated data.

Error variance was estimated using the residual sum of squares estimator proposed by **?** at a variety of tuning parameter values. This estimator has the following form:

$$\hat{\sigma}^2_{GL} = \frac{1}{n - \hat{s}_\lambda} \|Y - X\hat{\beta}_\lambda\|_2^2$$

where Y is the $n \times 1$ outcome vector, X is the design matrix based on the full model (1), λ is a value of the tuning parameter (specified on a grid in this study), $\hat{s}_\lambda$ is the number of nonzero entries in vector $\hat{\beta}_\lambda$, and $\hat{\beta}_\lambda$ is the regularization-based estimator of the coefficients. When the tuning parameter λ is relaxed to the extent that all coefficients are nonzero, the model becomes saturated and $\hat{\sigma}^2_{GL}$ is undefined. Cell means were estimated with the usual $X\hat{\beta}_\lambda$.

S2 Posterior model probability - additional results

The following rates report the proportion of times the flat intercept(flat additive) prior specification correctly identifies the true model. For additive data in setting 1: when $b = 4$, 0.902(0.949); when $b = 8$, 0.921(0.973); when $b = 12$, 0.951(0.981). For hidden additive data setting 2 with $\sigma^2 = 1$: when $b = 4$, 0.986(0.684); when $b = 8$, 0.997(0.986); when $b = 12$, 0.997(0.997). For setting 3 with $\sigma^2 = 5$: when $b = 4$, 0.531(0.121); when $b = 8$, 0.455(0.170); when $b = 12$, 0.362(0.180). For setting 4 when $\sigma^2 = 10$: when $b = 4$, 0.337(0.063), when $b = 8$, 0.188(0.038), when $b = 12$, 0.087(0.023).

The difference in posterior probability between the first most and second most probable models gives some indication of how much certainty the model selector recommends in a selected model. The following average differences indicate how much more probable the first model is compared with the second on average for the flat intercept(flat additive) prior in each setting. For additive data in setting 1: when $b = 4$, 0.536(0.696); when $b = 8$, 0.562(0.785); when $b = 12$, 0.576(0.842). For hidden additive data setting 2 with $\sigma^2 = 1$: when $b = 4$, 0.779(0.464); when $b = 8$, 0.961(0.909); when $b = 12$, 0.978(0.975). For setting 3 with $\sigma^2 = 5$: when $b = 4$, 0.396(0.613); when $b = 8$, 0.380(0.588); when $b = 12$, 0.320(0.524). For setting 4 when $\sigma^2 = 10$: when $b = 4$, 0.351(0.671), when $b = 8$, 0.299(0.701), when $b = 12$, 0.243(0.696). The difference between the top two models decreases when error variance increases, corresponding to less certainty in the model selector's recommendation when noise is high relative to signal in the data.

S3 Trace plots and diagnostics

Trace plots and Gelman-Rubin (**?**) upper bounds for the 199th (chosen randomly) Monte Carlo data set in each simulated setting. Top left panel in each plot is error variance. Other panels correspond to the cell means. Convergence of the Gibbs sampler is essentially immediate in all cases.

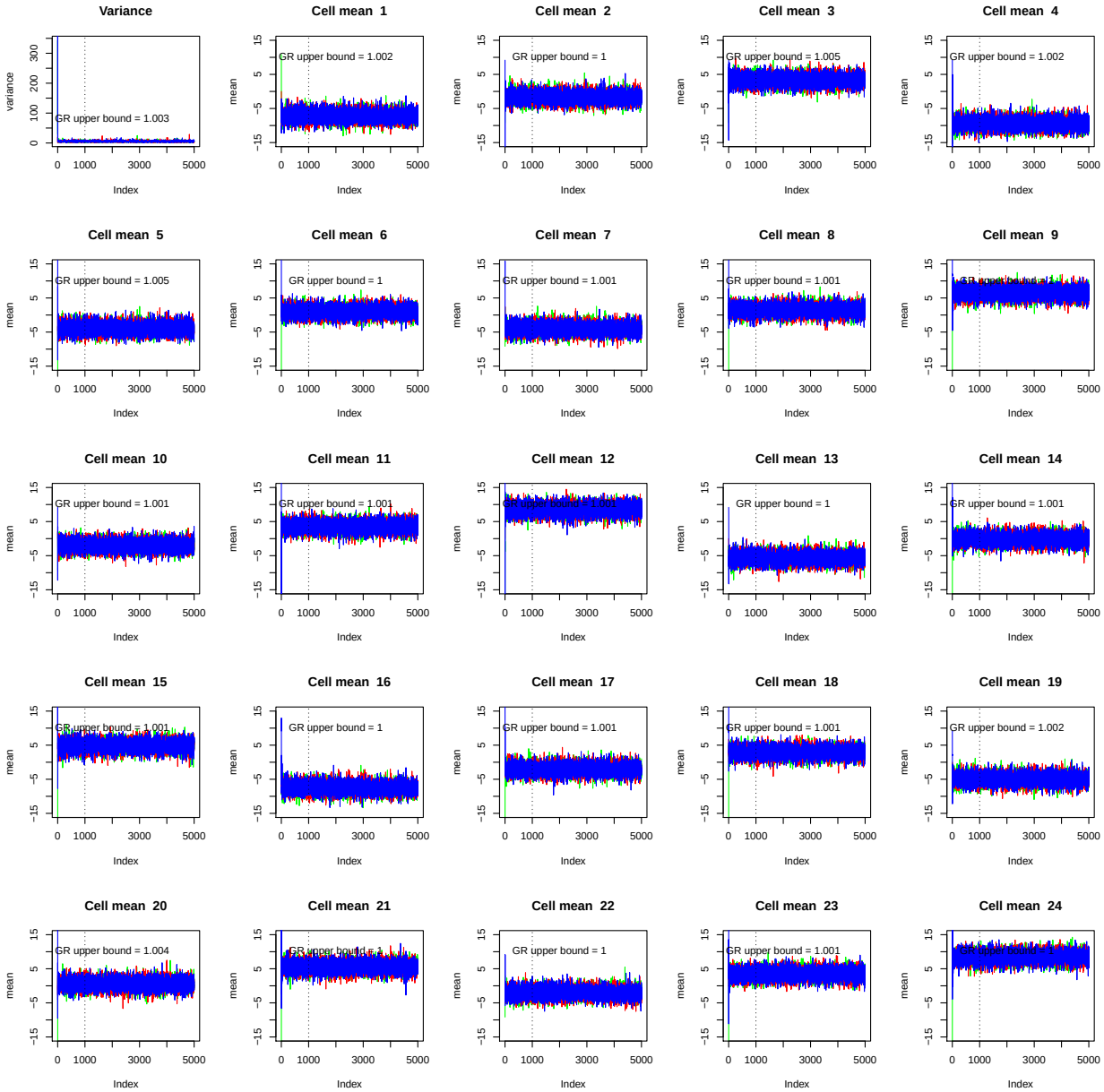

Figure S1: Setting 1 additive data. Trace plots correspond to Gibbs sampler for additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

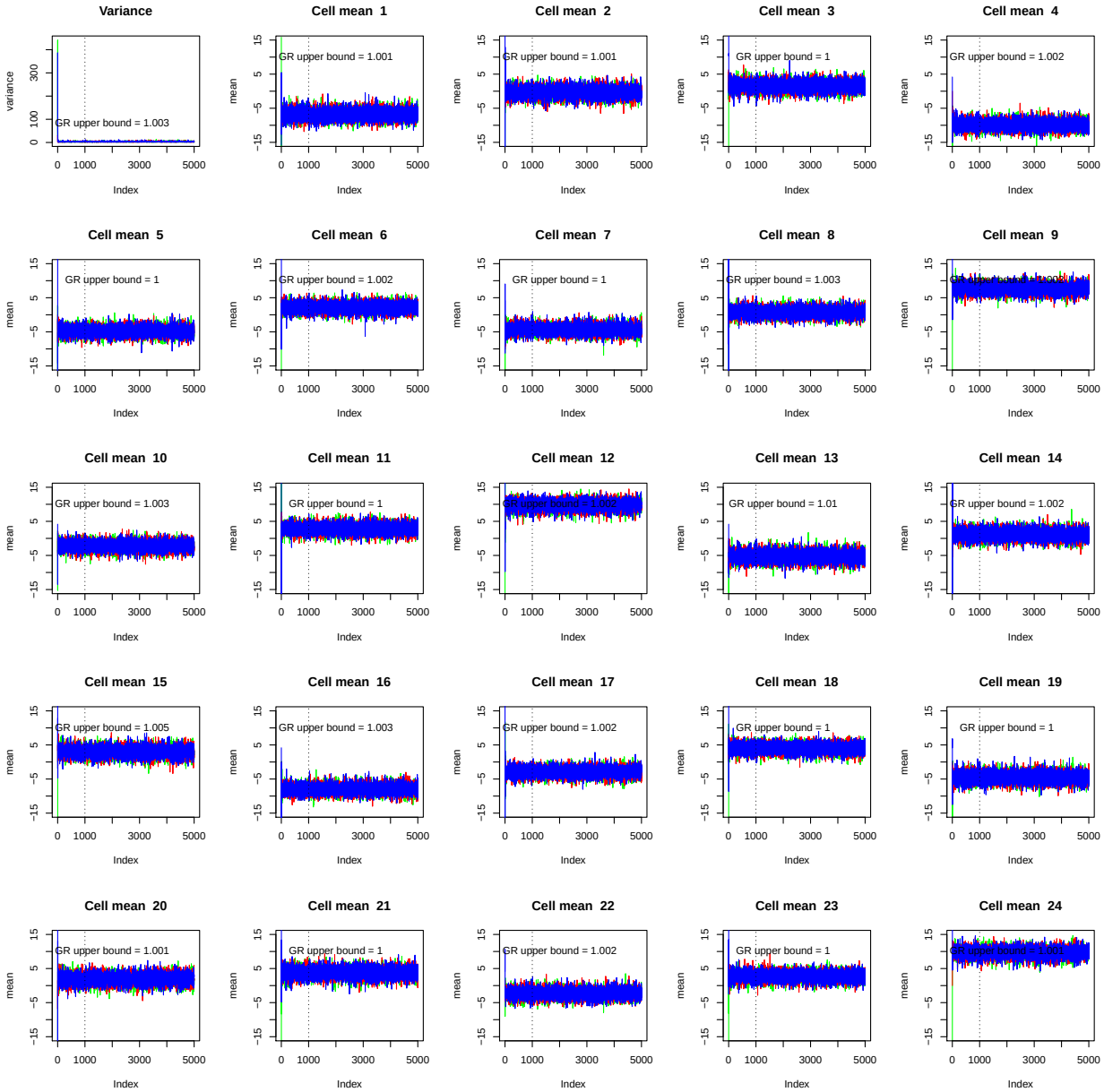

Figure S2: Setting 1 additive data. Trace plots correspond to Gibbs sampler for the most probable hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

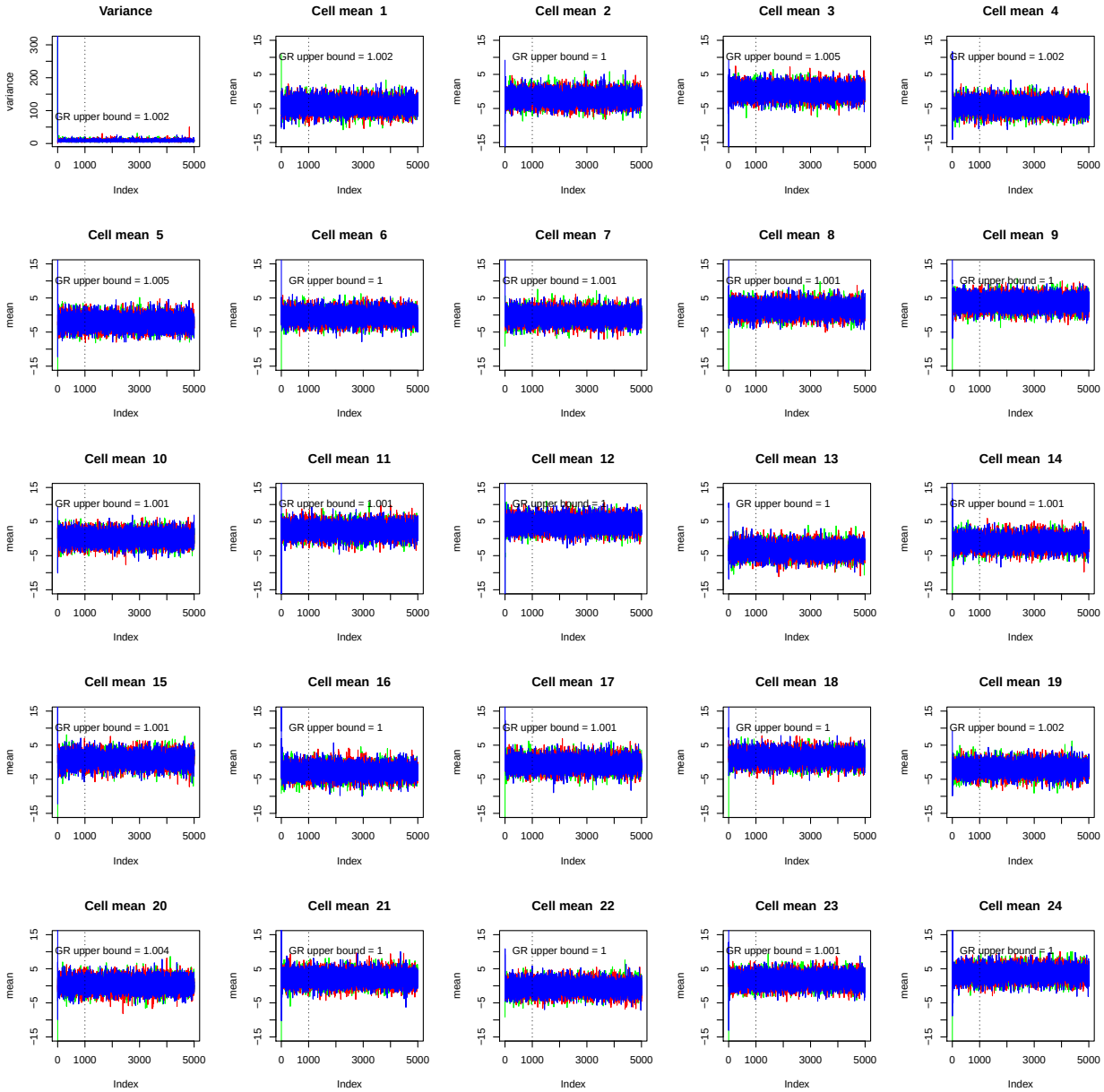

Figure S3: Setting 2 hidden additive data. Trace plots correspond to Gibbs sampler for additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

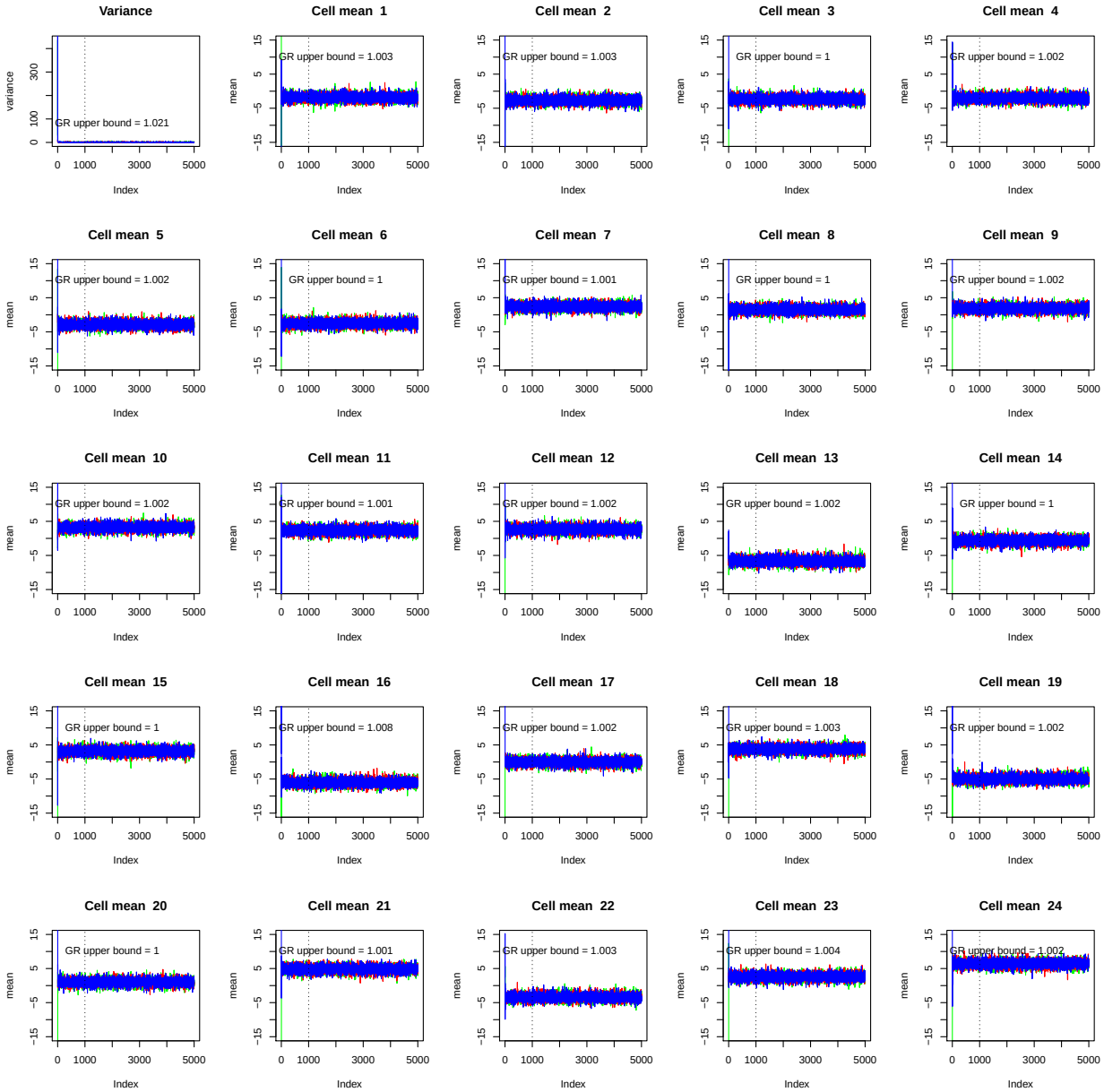

Figure S4: Setting 2 hidden additive data. Trace plots correspond to Gibbs sampler for most probable hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

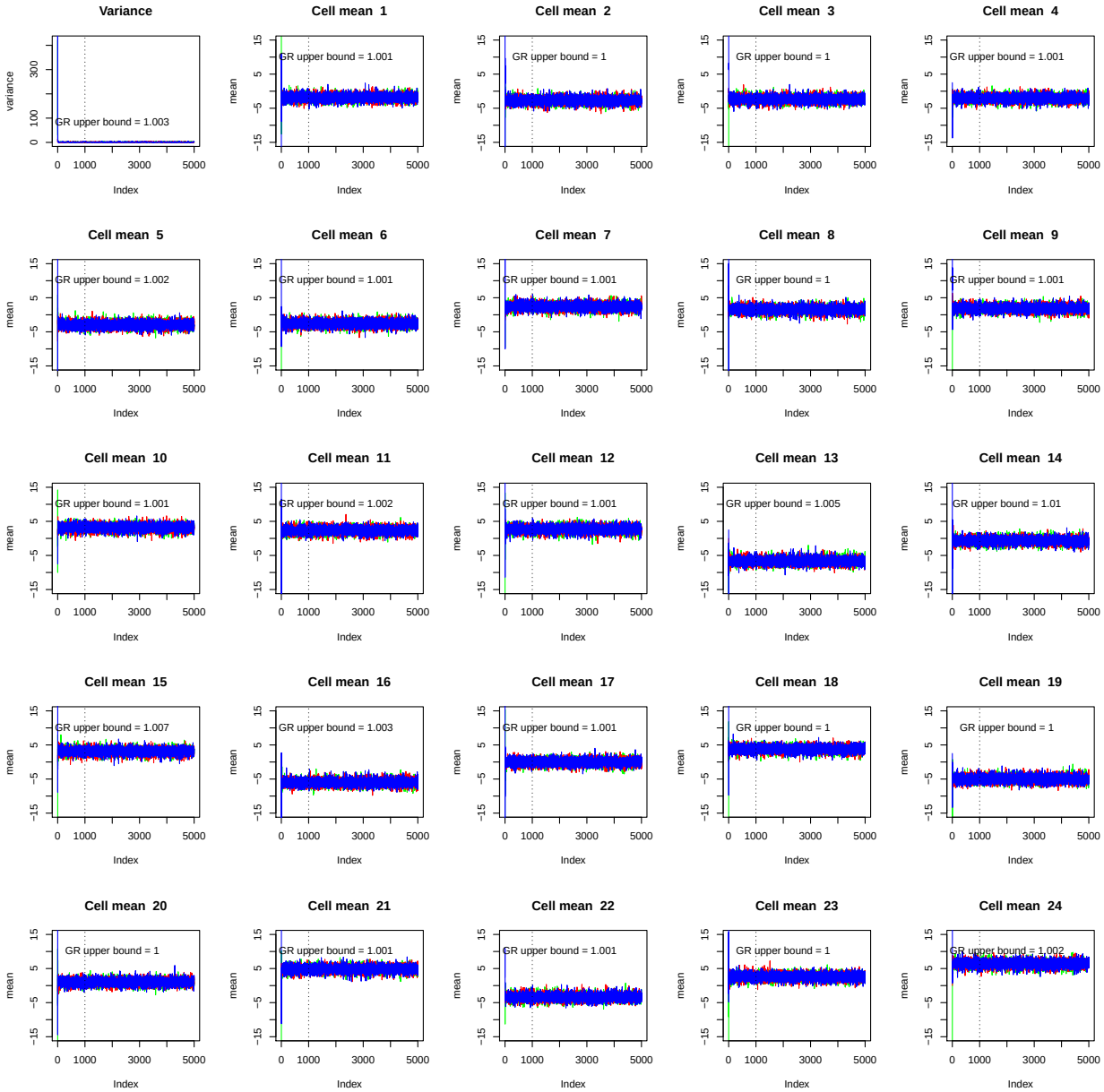

Figure S5: Setting 2 hidden additive data. Trace plots correspond to Gibbs sampler for true hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

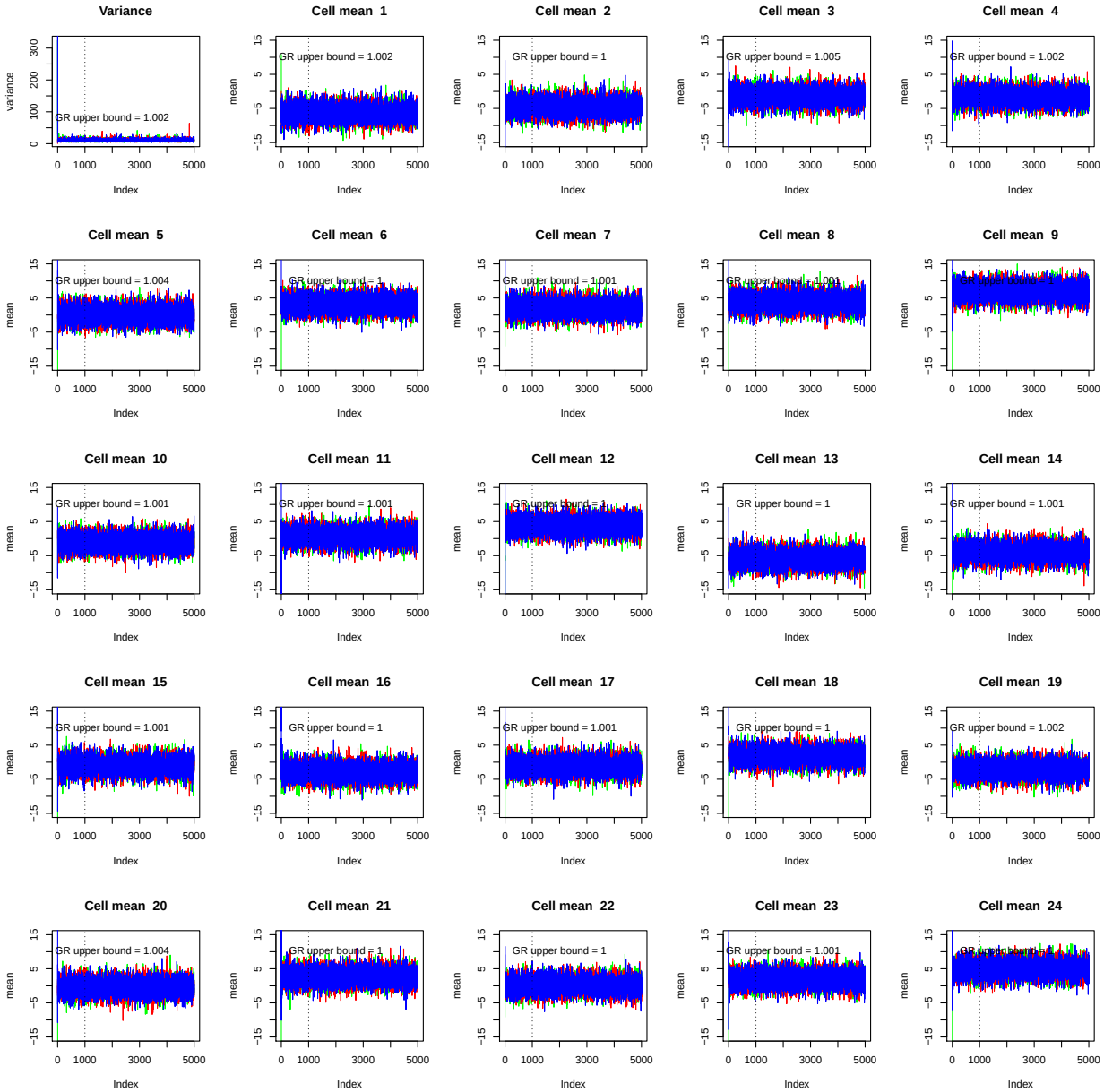

Figure S6: Setting 3 hidden additive data. Trace plots correspond to Gibbs sampler for additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

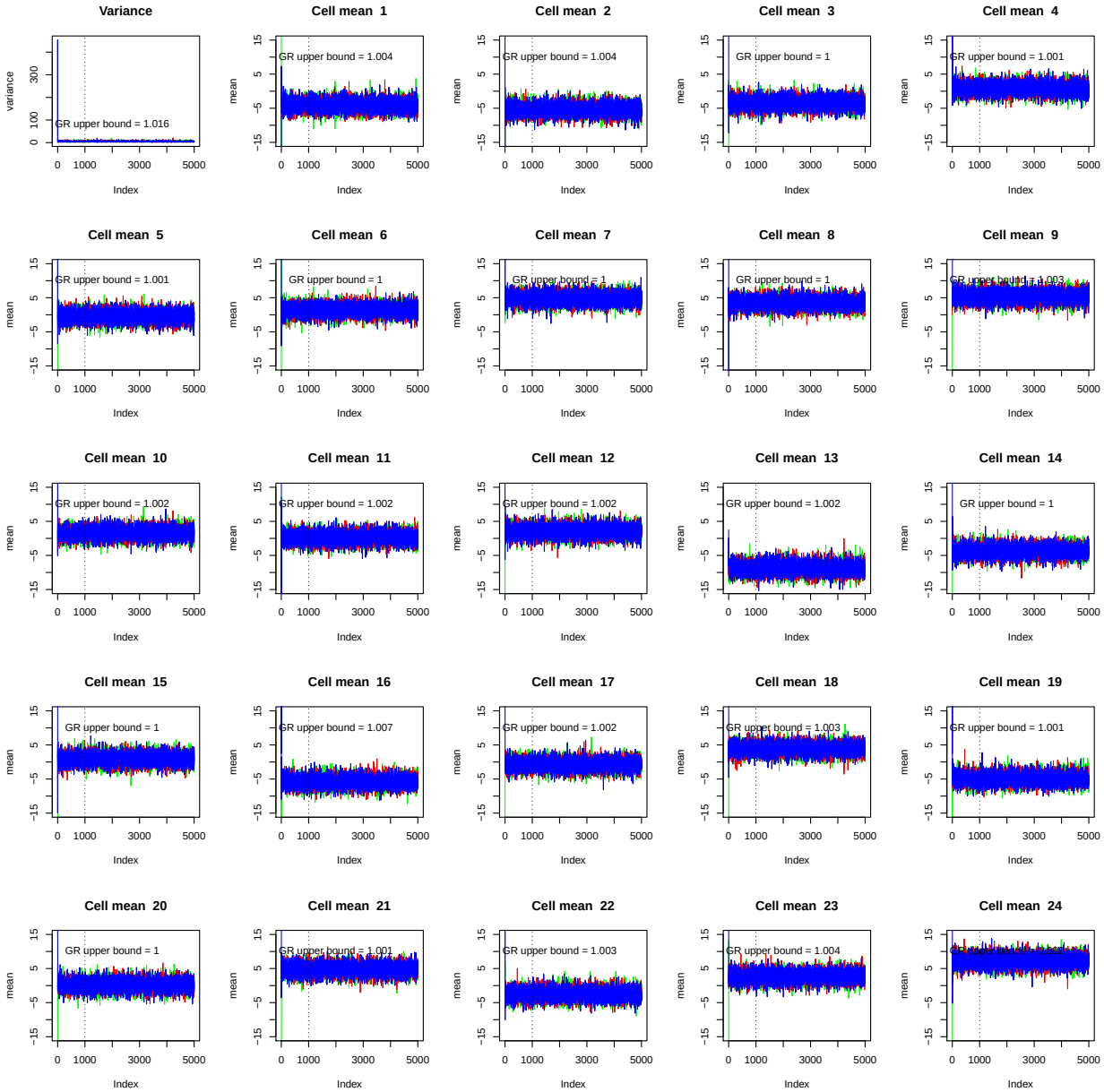

Figure S7: Setting 3 hidden additive data. Trace plots correspond to Gibbs sampler for most probable hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

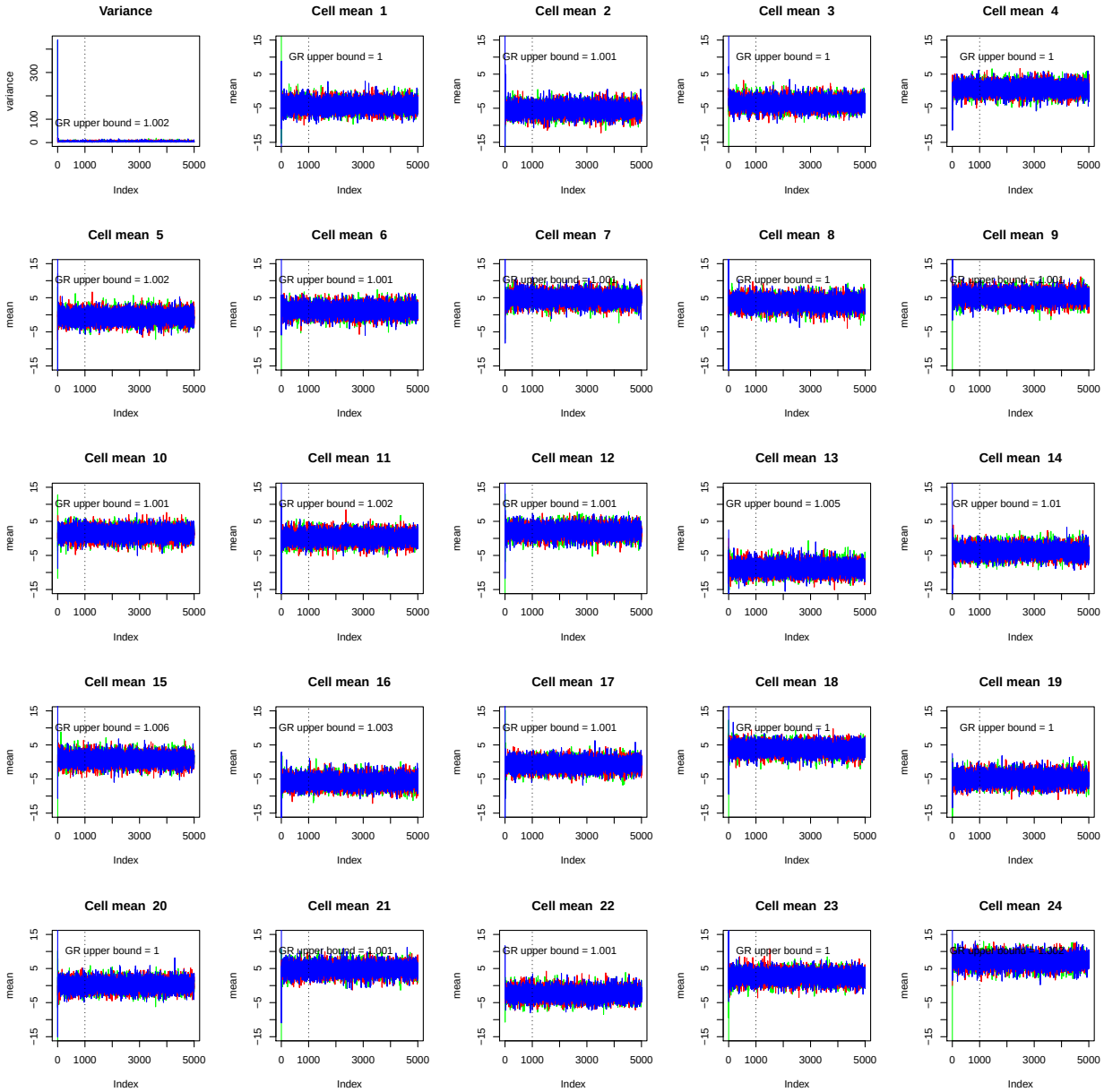

Figure S8: Setting 3 hidden additive data. Trace plots correspond to Gibbs sampler for true hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

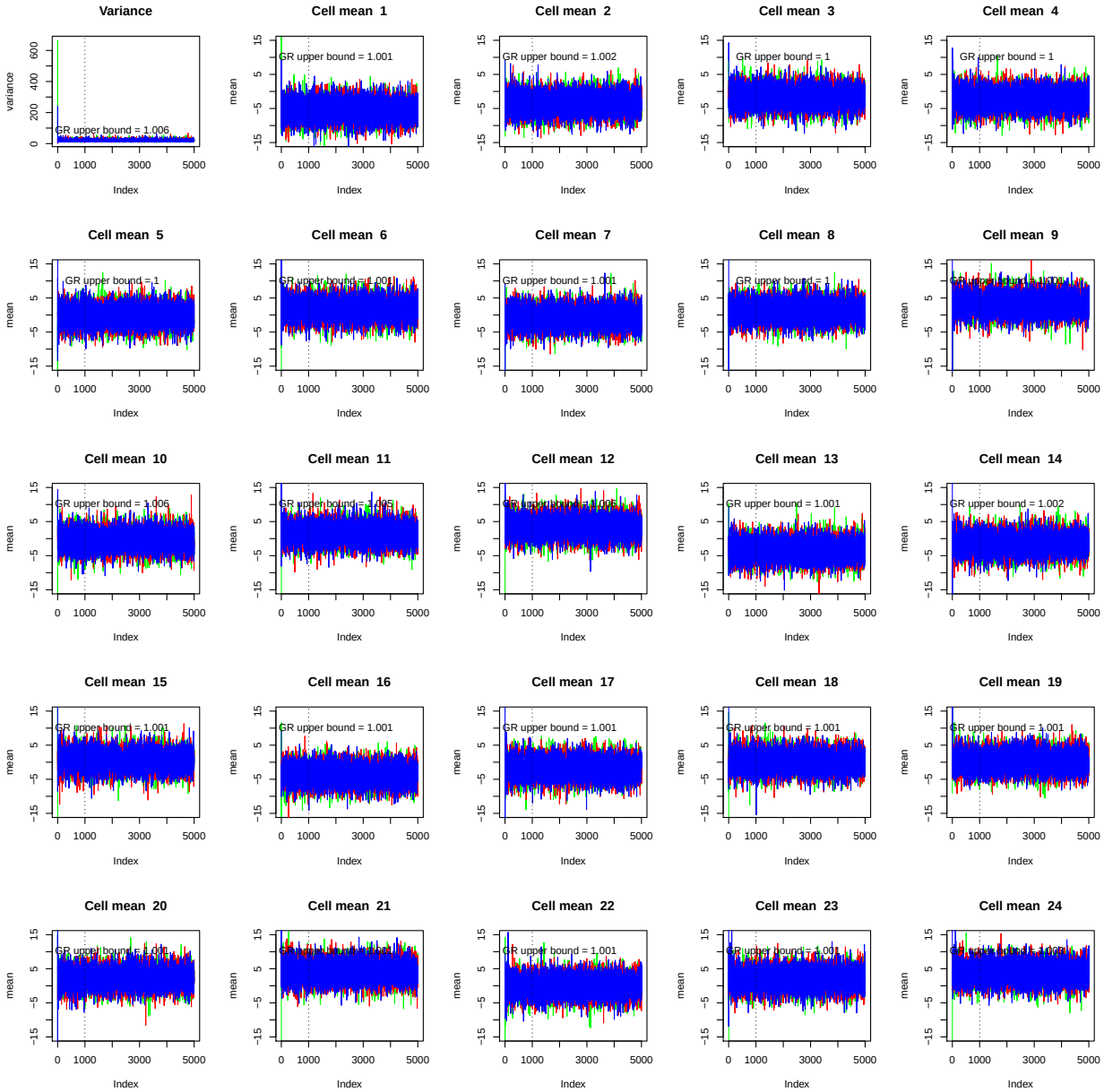

Figure S9: Setting 4 hidden additive data. Trace plots correspond to Gibbs sampler for additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

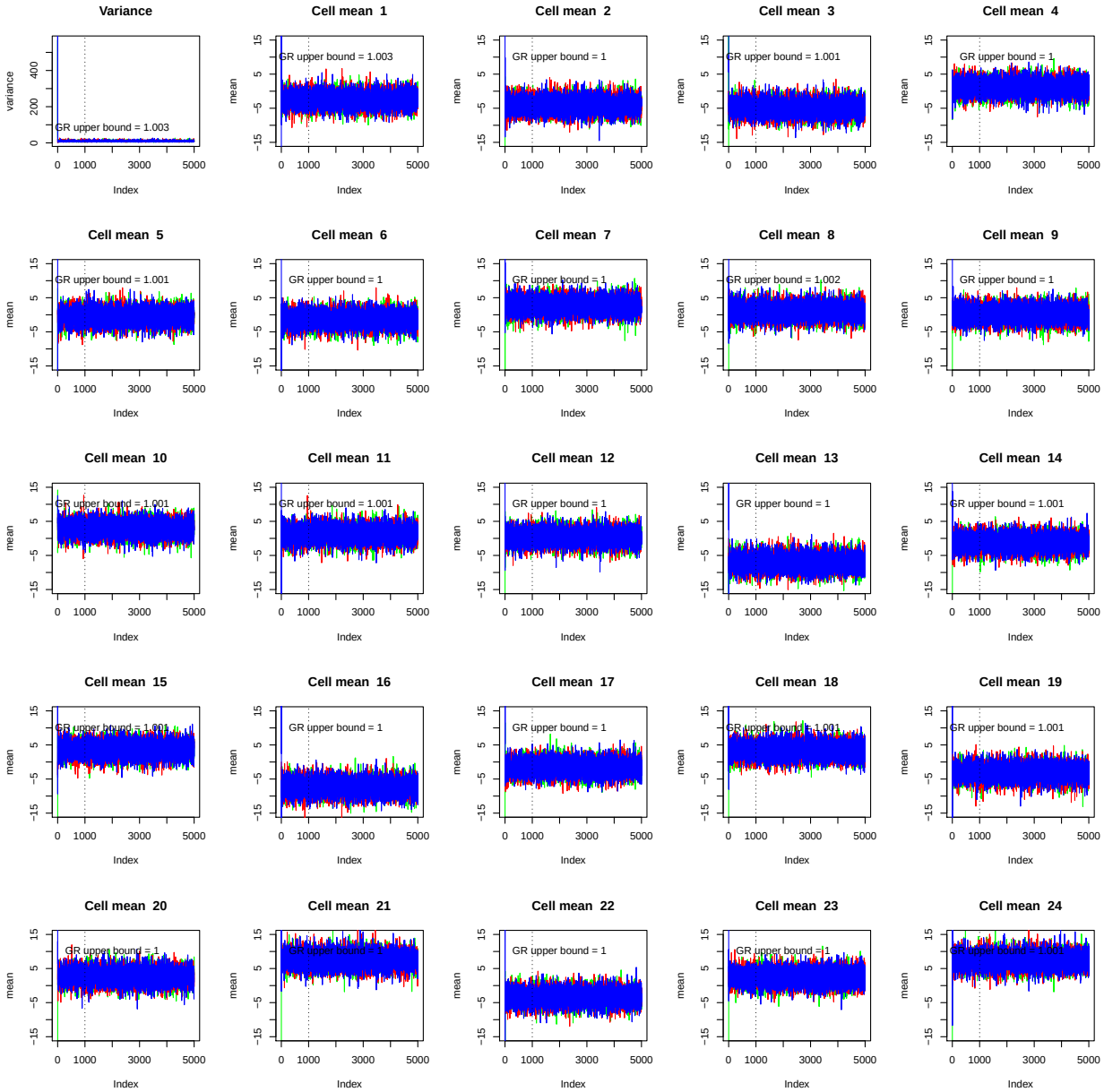

Figure S10: Setting 4 hidden additive data. Trace plots correspond to Gibbs sampler for most probable hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.

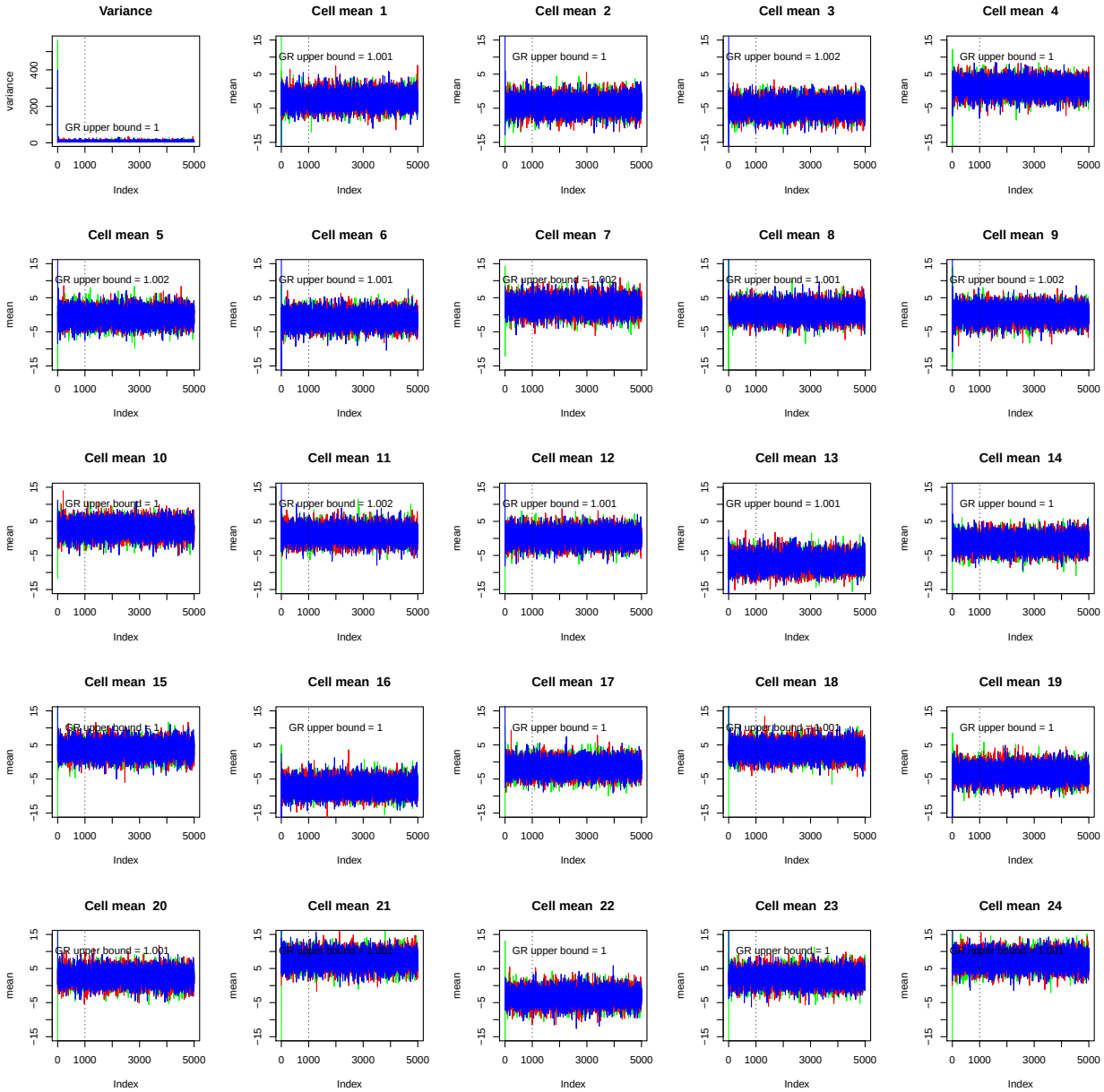

Figure S11: Setting 4 hidden additive data. Trace plots correspond to Gibbs sampler for true hidden additive model. Top left is error variance, other panels are for cell means. Gelman-Rubin upper bound superimposed on each panel indicates convergence.