

Robust Approximate Bayesian Inference with Synthetic Likelihood

David T. Frazier*

Christopher Drovandi†

June 12, 2020

Abstract

Bayesian synthetic likelihood (BSL) is now an established method for conducting approximate Bayesian inference in models where, due to the intractability of the likelihood function, exact Bayesian approaches are either infeasible or computationally too demanding. Implicit in the application of BSL is the assumption that the data generating process (DGP) can produce simulated summary statistics that capture the behaviour of the observed summary statistics. We demonstrate that if this compatibility between the actual and assumed DGP is not satisfied, i.e., if the model is misspecified, BSL can yield unreliable parameter inference. To circumvent this issue, we propose a new BSL approach that can detect the presence of model misspecification, and simultaneously deliver useful inferences even under significant model misspecification. Two simulated and two real data examples demonstrate the performance of this new approach to BSL, and document its superior accuracy over standard BSL when the assumed model is misspecified.

Keywords: approximate Bayesian computation; synthetic likelihood; likelihood-free inference; model misspecification; robust Bayesian inference; slice sampling.

1 Introduction

In situations where the likelihood of the underlying model is intractable, approximate Bayesian methods are often the only feasible solution to conduct Bayesian inference. Indeed, approximate Bayesian methods are an increasingly common tool in the arsenal of the

*Department of Econometrics and Business Statistics, Monash University, and the Australian Centre of Excellence for Mathematical and Statistical Frontiers

†School of Mathematical Sciences, Queensland University of Technology and the Australian Centre of Excellence for Mathematical and Statistical Frontiers

practicing statistician and allow users to conduct reliable inference in models where exact Bayesian inference procedures are either infeasible, or too computationally demanding.

The literature on approximate Bayesian inference now includes several competing approximate methods that are often useful in different scenarios. Arguably, the two most common likelihood-free Bayesian methods in the statistical literature are approximate Bayesian computation (ABC) (see, e.g., Marin et al., 2012 for a review) and Bayesian synthetic likelihood (BSL) (Wood, 2010, Price et al., 2018). The machine learning community is also making significant contributions to likelihood-free methods, such as using emulation to reduce the number of calls to the model simulator (e.g. Gutmann and Corander, 2016) and training neural conditional density estimators such as normalizing flows (e.g. Papamakarios et al., 2018). We refer to Cranmer et al. (2019) for a comprehensive review of machine learning approaches to likelihood-free methods.

Following the frequentest synthetic likelihood approach of Wood (2010), Price et al. (2018) develop an alternative to ABC by constructing a Bayesian version of synthetic likelihood, which places a prior distribution over the parameters and generates an approximate posterior. Unlike ABC, which implicitly estimates a version of the likelihood for the summaries, BSL directly assumes that the joint density of the summary statistics, conditional on the unknown model parameters, is Gaussian with unknown mean and variance. Using independent simulations obtained from the assumed data generating process (DGP), the mean and variance of the summary statistics are then estimated, and used to construct a (simulated) Gaussian likelihood function that is directly inserted into standard Markov Chain Monte Carlo (MCMC) algorithms. Price et al. (2018) demonstrate the BSL approach across several examples, and show that it often performs well in comparison with ABC.

BSL, and approximate Bayes methods more generally, are most often applied in situations where the complexity of the model that is assumed to have generated the observed data renders exact Bayesian inference infeasible. That is, by the very nature of the problems to which BSL is commonly applied, the model is so complex that we can not easily

access the DGP and must instead resort to an approximate inference approach. However, while complicated, highly-structured models that allow for vast complexity allow us to explain critical features of the observed data, it is unlikely that any modeler will be able to construct an entirely accurate model that captures all features of the observed data. In short, all models are wrong and the scientist cannot obtain a “correct” one through excessive elaboration (Box, 1976).

The implications of such a statement are particularly worrying in the context of BSL, where the assumed underlying DGP is often very complex. Indeed, applying the above reasoning of Box, it must be the case that the models to which BSL is routinely applied are misspecified representations of the actual, or “true”, DGP. In such situations, the application of BSL deserves further scrutiny given the recent results of Frazier et al. (2020), which demonstrate that if the model is misspecified the ABC posterior can be ill-behaved. Given that the principles underlying ABC and BSL are qualitatively the same, further analysis is needed to ensure that BSL does not suffer from the same issues as ABC in cases where the model is misspecified.

Through several simulated and empirical examples, we demonstrate that if the assumed model is misspecified, point estimators and credible sets obtained from BSL are unreliable. To circumvent this issue, we propose two novel versions of BSL that deliver “robust” inferences regardless of whether the model is correctly specified. Herein, we follow the robust statistics literature, as described in, e.g., Hampel et al. (2011), and consider a statistical inference procedure to be robust if it is not overly “sensitive” to departures from the underlying modeling assumptions.¹

This new BSL approach has three principle benefits over standard BSL. Firstly, this new procedure is less sensitive to model misspecification than standard BSL. In particular, the resulting posteriors are less affected by model misspecification than those obtained

¹Formalizing this notion of robustness requires a refined mathematical treatment that is beyond the scope of this paper. A precise definition of robustness requires the specification and use of infinitesimal neighborhoods that capture the degree of model misspecification, and which then allow us to formally define a notion “sensitivity” (see, e.g., Hampel et al., 2011, and Rieder, 2012 for a discussion). Given this, we leave a formal study on the theoretical robustness of this new BSL approach for future research.

from BSL (see Section 3.3 for a detailed discussion and Section 4 for specific examples). Consequently, this approach yields more reliable point estimators and uncertainty quantification in misspecified models. Second, this new BSL approach has an in-built mechanism for diagnosing model misspecification, which allows us to discern which components of the model may in-fact be misspecified. Lastly, this new approach is computationally robust in the following sense: when the model is misspecified, the standard BSL posterior can require an excessive number of model simulations to generate accurate samples, however, our proposed BSL approach does not suffer from this issue. Given the above notions of robustness, both statistical and computational, we refer to this new approach as robust BSL (R-BSL).

Through a series of examples, both simulated and empirical, we demonstrate that R-BSL yields reliable statistical inferences regardless of whether the model is correctly or incorrectly specified, and can consistently detect when the modeling assumptions are violated. Repeated sampling results demonstrate that R-BSL yields more accurate point estimators and quantifies uncertainty better than BSL in misspecified models. In addition, theoretical (and simulation results) demonstrate that if the model is correctly specified, R-BSL and BSL deliver similar statistical inferences. As such, R-BSL allows users to hedge against model misspecification, but ensures that the resulting inference remains accurate if the model is correctly specified.

Before moving on, we remark that throughout this paper, we assume that the Gaussian assumption of the synthetic likelihood is at least approximately correct. We refer the reader to the discussion in Section 5 for more details, and potential impacts on the resulting inferences when this assumption is invalid.

The remainder of the paper is organized as follows. In Section two we give a brief overview of BSL and examine the consequences of model misspecification in BSL. Section three presents our robust approach to BSL. Through a sequence of examples discussed in Section four, we document the good performance of R-BSL and the poor performance of BSL across models with varying levels of misspecification. In addition, Section four

contains an empirical application to the analysis of invasive species. Section five concludes. Additional examples and the proofs of the technical results are given in the supplementary material.

2 Bayesian Synthetic Likelihood and Compatibility

2.1 Bayesian Synthetic Likelihood Framework

We observe data $\mathbf{y} = (y_1, \dots, y_n)^\top$, $n \geq 1$, and denote by P_0^n the true distribution of the observed sample. The true distribution is unknown and instead we consider that the class of probability measures $\{\theta \in \Theta \subset \mathbb{R}^{d_\theta}, n \geq 1 : P_\theta^n\}$, for some value of θ , have generated the data, and denote the corresponding conditional density as $p_n(\cdot|\theta)$. Given prior beliefs over the unknown parameters in the model θ , represented by the probability measure $\Pi(\theta)$, with its density denoted by $\pi(\theta)$, our aim is to produce draws from the exact posterior density

$$\pi(\theta | \mathbf{y}) \propto p_n(\mathbf{y}|\theta)\pi(\theta).$$

In situations where the likelihood is intractable, sampling from $\pi(\theta | \mathbf{y})$ can be computationally costly or infeasible, however, so-called likelihood-free methods can still be used to conduct inference on the unknown parameters θ . The most common implementations of these methods in the statistical literature are approximate Bayesian computation (ABC) and Bayesian synthetic likelihood (BSL). Both ABC and BSL generally degrade the data down to a vector of summary statistics and then perform posterior inference on the unknown θ , conditional only on this vector of summary statistics.

More formally, let $\eta(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{d_\eta}$ denote a d_η -dimensional map, $d_\eta \geq d_\theta$, that represents the chosen summary statistics, and let $\mathbf{z} := (z_1, \dots, z_n)^\top \sim P_\theta^n$ denote data simulated from the model P_θ^n . For $G_n(\cdot|\theta)$ denoting the projection of P_θ^n under $\eta(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{d_\eta}$, with $g_n(\cdot|\theta)$ its corresponding density, the goal of approximate Bayesian methods is to generate

samples from the approximate or ‘partial’ posterior

$$\pi[\theta \mid \eta(\mathbf{y})] \propto g_n[\eta(\mathbf{y}) \mid \theta] \pi(\theta).$$

However, given the complexity of the assumed model, P_θ^n , it is unlikely that the structure of $G_n(\cdot \mid \theta)$ is any more tractable than the original likelihood function $p_n(\mathbf{y} \mid \theta)$. Therefore, simulation-based sampling schemes must be applied to generate samples from $\pi[\theta \mid \eta(\mathbf{y})]$.

The approximate methods of ABC and BSL differ in how $g_n(\cdot \mid \theta)$ is estimated. ABC forms an implicit nonparametric estimator of $g_n(\cdot \mid \theta)$, while BSL uses a parametric or semi-parametric (An et al., 2020) approximation of $g_n(\cdot \mid \theta)$. In particular, BSL replaces - an estimate of - the (intractable) density $g_n(\cdot \mid \theta)$ by a multivariate Gaussian approximation:

$$\mathcal{N}[\eta; \mu(\theta), \Sigma(\theta)],$$

where $\mu(\theta)$ and $\Sigma(\theta)$ denote the mean and variance of the summary statistics. In cases where $\mu(\theta), \Sigma(\theta)$ are known we can obtain the “exact” BSL posterior

$$\pi[\theta \mid \eta(\mathbf{y})] \propto \mathcal{N}[\eta(\mathbf{y}); \mu(\theta), \Sigma(\theta)] \pi(\theta).$$

However, in almost any practical example $\mu(\theta)$ and $\Sigma(\theta)$ are unknown and we must replace these quantities with the estimated counterparts $\mu_m(\theta)$ and $\Sigma_m(\theta)$, obtained as

$$\mu_m(\theta) = \frac{1}{m} \sum_{i=1}^m \eta(\mathbf{z}^i), \quad \Sigma_m(\theta) = \frac{1}{m} \sum_{i=1}^m [\eta(\mathbf{z}^i) - \mu_m(\theta)] [\eta(\mathbf{z}^i) - \mu_m(\theta)]^\top,$$

and where each simulated data set \mathbf{z}^i , $i = 1, \dots, m$, are generated iid from P_θ^n ; that is, both $\mu_m(\theta)$ and $\Sigma_m(\theta)$ depend on the simulated data sets $\mathbf{z}^1, \dots, \mathbf{z}^m$. The Gaussian approximation is then directly used within an MCMC sampling scheme to sample from the

following approximation to the partial posterior, hereafter referred to as the BSL posterior,

$$\begin{aligned} \hat{\pi}[\theta \mid \eta(\mathbf{y})] &\propto \bar{g}_n[\eta(\mathbf{y}) \mid \theta] \pi(\theta), \\ \bar{g}_n[\eta(\mathbf{y}) \mid \theta] &:= \int \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta), \Sigma_m(\theta)] \left\{ \prod_{i=1}^m g_n[\eta(\mathbf{z}^i) \mid \theta] \right\} d\eta(\mathbf{z}^1) \cdots d\eta(\mathbf{z}^m). \end{aligned} \tag{1}$$

Price et al. (2018) demonstrate empirically that the BSL posterior depends weakly on m , provided that m is chosen large enough so that the plug-in synthetic likelihood estimator has a small enough variance to ensure that MCMC mixing is not adversely affected. In this paper we choose m and the number of MCMC iterations large enough so that the Monte Carlo error arising from MCMC is small.

Due to the parametric nature of equation (1), BSL can often treat summary statistics of larger dimension than ABC and can lead to sharper inference in some cases. While the validity of the Gaussian approximation is often warranted if the underlying summaries satisfy a central limit theorem (Wood, 2010), even in cases where the summary statistics are far from Gaussian, BSL has displayed some insensitivity to violations of this assumption (Price et al., 2018). However, if the statistics are very far from being Gaussian, this can result in a significant loss of accuracy, see An et al. (2020) for a demonstration.

2.2 Model Incompatibility and its Consequences

BSL implicitly maintains that the assumed model can generate simulated summary statistics $\eta(\mathbf{z})$ that can match the observed summary statistics $\eta(\mathbf{y})$. That is, BSL is not required to match every aspect of the data, but only those features of the data that are captured via the summary statistics $\eta(\mathbf{y})$. This differs from a standard Bayesian framework based on a likelihood, where, under general regularity conditions, the posterior ultimately gives higher probability mass to values of $\theta \in \Theta$ that ensure the Kullback-Leibler (KL) divergence

$$\mathcal{D}(P_0^n \parallel P_\theta^n) = \int \log \left\{ \frac{p_0^n(\mathbf{y})}{p_n(\mathbf{y} \mid \theta)} \right\} dP_0^n(\mathbf{y})$$

is as close to zero as possible. When the model is misspecified, i.e., when $P_0^n \neq P_\theta^n$ for any $\theta \in \Theta$, following Kleijn and Van der Vaart (2012), the posterior eventually places increasing mass on the value $\theta^* \in \Theta$ that minimizes the KL-divergence.

Given that BSL attempts to simulate summary statistics $\eta(\mathbf{z})$ that can match the value of the observed summary statistic $\eta(\mathbf{y})$, as measured by a weighted Euclidean norm, KL divergence is not the most meaningful notion of model misspecification associated with BSL. A more meaningful notion is whether or not $\eta(\mathbf{z})$ can match $\eta(\mathbf{y})$ in terms of the Euclidean norm. Therefore, we follow Marin et al. (2014), and Frazier et al. (2020), and say that the assumed model is misspecified when it can not generate summaries $\eta(\mathbf{z})$ that can match $\eta(\mathbf{y})$. More formally, for $b_0 := \text{plim}_n \eta(\mathbf{y})$ and $b(\theta) := \text{plim}_n \eta(\mathbf{z})$, denoting the probability limits of the summaries as $n \rightarrow \infty$, we can state this notion of misspecification as follows.²

Definition 1. *The model $P_\theta^n \times \Pi$ and summary statistic map $\eta(\cdot)$ are compatible if*

$$\inf_{\theta \in \Theta} \|b(\theta) - b_0\| = 0.$$

Heuristically, compatibility requires that asymptotically $\eta(\mathbf{y})$ must be in the range of $\eta(\mathbf{z})$ and implies that, for some value of θ , $\eta(\mathbf{z})$ can recover $\eta(\mathbf{y})$ when \mathbf{z} is simulated under P_θ^n . Compatibility is not concerned with the distributions of $\eta(\mathbf{y})$ and $\eta(\mathbf{z})|\theta$, which ultimately must be degenerate if they are to be informative about θ , but only their probability limits. Employing this notion of model misspecification allows us to analyze a large set of examples, since all this concept requires is that the summaries satisfy some weak law of large numbers.³

When the model is not compatible, it can not (asymptotically) replicate the value of the observed statistics and, following the nomenclature in Frazier et al. (2020), we say that

²The fact that we require the summary statistics in BSL to concentrate to well-defined limit counterparts should not come as a surprise. Frazier et al. (2018) have demonstrated that this concentration is necessary to formally discuss the asymptotic behavior of ABC, while Frazier et al. (2019) echos this finding in the specific context of BSL.

³It is also useful to point out that compatibility is not directly related to, or interpreted in terms of, any statistical divergence, and is only related to the minimum achievable distance between observed and simulated summaries (in the infinite data limit). While it may be possible to recast compatibility in terms of a statistical divergence, it is not clear what, if any, additional insights such an analysis would yield.

the model is *misspecified in the BSL sense*.⁴ As recently discussed by Frazier et al. (2020), in the context of ABC, when approximate methods are based on a model and summary statistic combination that are not compatible, the resulting posteriors can be ill-behaved and statistical inferences based on these posteriors can be highly-unreliable. Given that ABC and BSL are based on the same principles, it is highly likely that BSL will suffer from the same issues as ABC when the above compatibility condition is not satisfied. While we demonstrate this with realistic examples in Section 4, we first consider an artificially simple example where BSL should perform well, but due to model misspecification, BSL inference can be unreliable.

2.3 Toy Example 1: Contaminated Normal Model

To demonstrate how BSL can fail under model misspecification, we consider an artificially simple example: our goal is inference on the unknown mean parameter θ in the assumed model

$$y_i = \theta + v_i, \quad v_i \stackrel{iid}{\sim} \mathcal{N}(0, 1). \quad (2)$$

However, we consider that the assumed model in (2) is correct, but only for a portion of the data, $\omega \in (0, 1)$, while the remaining portion of the data, $1 - \omega$, is contaminated by outliers that are also normal but can have a much larger variance than unity. The true DGP can then be stated as the mixture model

$$y_i = \begin{cases} \theta + \epsilon_{1,i}, & \epsilon_{1,i} \sim \mathcal{N}(0, 1), \quad \text{with probability } \omega \\ \theta + \epsilon_{2,i}, & \epsilon_{2,i} \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad \text{with probability } 1 - \omega \end{cases}. \quad (3)$$

Given that our assumed model is (2), the most reasonable set of summary statistics to choose are the sample mean $\eta_1(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^n y_i$ and sample variance $\eta_2(\mathbf{y}) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \eta_1(\mathbf{y}))^2$. It can easily be shown that when σ_ϵ^2 in the true DGP, (3), satisfies $\sigma_\epsilon^2 \neq 1$, the

⁴This notion of model misspecification is a “global” notion of misspecification, and is precisely the same notion of model misspecification used in Marin et al. (2014) and Frazier et al. (2020). This is in contrast to the notion of “local” misspecification that is commonly entertained in the robust statistics literature.

model is not compatible with the variance summary, and Definition 1 is not satisfied.

Regardless of this model misspecification, one would think that BSL should perform well: the first summary statistic, $\eta_1(\mathbf{y})$, is Gaussian, the statistic $\eta_2(\mathbf{y})$ satisfies the central limit theorem, and $\eta_1(\mathbf{y})$ is sufficient for θ . Moreover, the model is only wrong for a proportion of the data.

For this experiment, we fix the level of data contamination at $1 - \omega = 0.20$, and generate simulated data sets for \mathbf{y} where each data set corresponds to a different value of σ_ϵ^2 . We choose a grid of values for σ_ϵ^2 so that the *sample (and population) standard deviation* of \mathbf{y} range from 1 to 2, with evenly spaced increments of 0.10. The observed data is generated so that the *sample (and population) mean* is fixed at 1 for all simulated data sets.⁵ The sample size across the experiments is taken to be $n = 100$. Our prior for θ is $\theta \sim \mathcal{N}(0, 10)$.

For BSL, we use $m = 100$ simulated data sets to estimate the mean and variance of the summaries. Sampling of the BSL posterior is implemented using random-walk Metropolis-Hastings (RWMH), where the variance of the proposal is set to the exact posterior variance. The sampler is initialized at $\theta = 0$, and run for 25,000 iterations, with the first 10,000 iterations discarded for burn-in.

Panel A of Figure 1 plots the acceptance rates across the different levels of model misspecification. From the results in Panel A, we see that at large levels of misspecification RWMH has a very difficult time exploring the parameter space. As a consequence, the MCMC chain stays stuck for long periods of time and the acceptance rates plummet from a peak of about 70%, to a low of less than 0.01%. In Panel B of Figure 1 we see that the posterior median of BSL, and the resulting credible set, vary significantly as the level of model misspecification in the data increases. Consequently, for values of the sample standard deviation larger than about 1.60, statistical inference based on the BSL posterior is not reliable.

The inaccuracy with which the BSL posterior is sampled is a consequence of the in-

⁵Since all BSL observes for the purpose of inference on θ is the sample mean and variance, generating data so that the sample mean and variance take on fixed values allows us to isolate the impact of model misspecification.

compatibility for $\eta_2(\mathbf{y})$: no matter the value of θ , the summary statistic $\eta_2(\mathbf{y})$ can not be matched by the simulated counterpart $\eta_2(\mathbf{z})$. Therefore, the actual value of the observed statistic $\eta_2(\mathbf{y})$, for any value of θ , will be in the tails of the Gaussian approximation for the posterior, which are inherently much noisier than values in the central mass of the approximation. Subsequently, this leads to a very noisy acceptance step and causes the MCMC chain to stick, with the overall result being unreliable inference for θ .

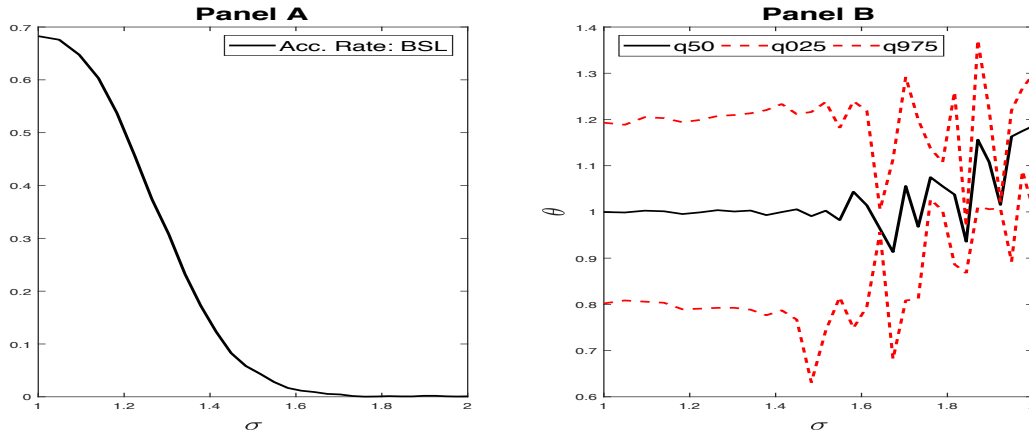


Figure 1: Panel A: acceptance rates for BSL across all levels of misspecification. Panel B: posterior medians and credible sets for BSL across the different levels of model misspecification. In both panels, the value of σ denotes the sample (and population) standard deviation for the experiment.

3 Robust Bayesian Synthetic Likelihood

We propose two possible strategies for conducting inference using BSL when the model and summaries are incompatible (i.e., when the model is misspecified in the BSL sense). The first approach augments the mean of the simulated summaries with additional free parameters, while the second approach augments the variance of the simulated summaries with additional free parameters. Both specifications allow us to conduct reliable statistical inference on the model parameters regardless of model misspecification, and allows us to determine which of the summaries are incompatible with the data. Given this robustness, throughout the remainder we refer to these approaches as “mean” and “variance” robust

BSL (R-BSL).⁶

3.1 Mean Adjustment and Prior Specification

Incompatibility implies that the observed statistics $\eta(\mathbf{y})$ can not be recovered by the simulated mean $\mu_m(\theta)$, for any $\theta \in \Theta$, with probability converging to one.⁷ Therefore, one approach to create a BSL procedure that is robust to this incompatibility issue is to **adjust the vector of simulated means**. This can be accomplished by **adding to $\mu_m(\theta)$ an additional free parameter Γ , where $\Gamma = (\gamma_1, \dots, \gamma_{d_\eta})^\top \in \mathcal{G} \subset \mathbb{R}^{d_\eta}$, so that $\eta(\mathbf{y})$ will always be in the support of this new simulated mean, even as the sample size diverges.** Defining the joint vector of unknown parameters as $\zeta := (\theta^\top, \Gamma^\top)^\top \in \Theta \times \mathcal{G} \subset \mathbb{R}^{d_\theta} \times \mathbb{R}^{d_\eta}$, **we define the vector of simulated means for use in BSL as**

$$\phi_m(\zeta) = \mu_m(\theta) + \text{diag}[\Sigma_m^{1/2}(\theta)] \Gamma.$$

Note that, by considering the scaled adjustment term, $\text{diag}[\Sigma_m^{1/2}(\theta)]\Gamma$, we ensure that these components are measured in the same units as $\mu_m(\theta)$, which allows us to treat Γ as if they were unitless.

Given this linear adjustment, and under weak conditions on the summary statistics and the parameter space $\Theta \times \mathcal{G}$, it is simple to see that $\phi_m(\zeta)$ will be compatible with $\eta(\mathbf{y})$ for any prior choice on Γ such that each individual component of Γ has support over \mathbb{R} .

Denote the prior on ζ by $\pi(\zeta)$. Following Price et al. (2018), **the augmented BSL target, which we refer to as the Robust BSL-mean (R-BSL-M) posterior, is to generate samples**

⁶We remind the reader that, following the robust statistics literature, we say a procedure is robust if it is not overly sensitive to model misspecification and/or if it can accurately diagnose model misspecification. Since this new approach accomplishes both of these tasks, the robust moniker is appropriate.

⁷Recall that $\eta(\mathbf{y})$ is most often a sample mean so that its support is, with large probability, a shrinking ball around the point b_0 .

from⁸

$$\hat{\pi}[\zeta \mid \eta(\mathbf{y})] \propto \bar{g}_n[\eta(\mathbf{y})|\zeta] \pi(\zeta),$$

where

$$\bar{g}_n[\eta(\mathbf{y})|\zeta] = \int \mathcal{N}[\eta(\mathbf{y}); \phi_m(\zeta), \Sigma_m(\theta)] \left\{ \prod_{i=1}^m g_n[\eta(\mathbf{z}^i) \mid \theta] \right\} d\eta(\mathbf{z}^1) \cdots d\eta(\mathbf{z}^m). \quad (4)$$

Prior Choice: Laplace Prior

To ensure that the observed summary $\eta(\mathbf{y})$ can always be recovered by $\phi_m(\zeta)$, even when n is very large, our prior on the components of Γ should allow for $\text{diag}[\Sigma_m^{1/2}(\theta)]\Gamma$ to escape the support of $\mu_m(\theta)$ with large probability. However, given that some components of the original $\mu_m(\theta)$ are likely compatible with some components of $\eta(\mathbf{y})$, we want to make sure that $\text{diag}[\Sigma_m^{1/2}(\theta)]\Gamma$ does not unduly perturb the components that are compatible. Therefore, we should choose a prior that places the vast majority of its mass near the origin. In this way, our prior choice for Γ should induce “*shrinkage*” in the components of Γ : only the components of Γ that correspond to incompatible summaries should receive significant posterior probability away from the origin, while the components of Γ that correspond to compatible summaries should have the majority of their posterior mass near the origin.

With these dual requirements in mind, and given that each component of Γ has the same prior scale, we propose to follow the Bayesian Lasso literature (Park and Casella, 2008) and use independent Laplace (i.e., double-exponential) priors for each component of

⁸The scaling of the perturbation by $\text{diag}[\Sigma_m^{1/2}(\theta)]$, gives the impression that the magnitude of the perturbation is decreasing as n increases, since $\text{diag}[\Sigma_m^{1/2}(\theta)]$ is decreasing as n increases. However, a simple exploration of the R-BSL-M posterior demonstrates that this scaling has no effect on these components, as they are themselves weighted by $\Sigma_m^{-1/2}(\theta)$ within the Gaussian kernel.

Γ , with fixed location 0 and common scale $\lambda > 0$:

$$\pi(\Gamma) := \prod_{j=1}^{d_\eta} \frac{1}{2\lambda} e^{-\frac{|\gamma_j|}{\lambda}} = \left(\frac{1}{2\lambda}\right)^{d_\eta} e^{-\frac{1}{\lambda} \sum_{j=1}^{d_\eta} |\gamma_j|}. \quad (5)$$

When convenient, we denote this prior by $\text{La}(0, \lambda)$. The Laplace prior for Γ guarantees that the majority of prior mass for γ_j is near the origin, but has thick enough tails so that $\phi_m(\zeta)$ is compatible with virtually any $\eta(\mathbf{y})$ that would be used in practice.

The hyper-parameter λ should be chosen so that the prior support of $\pi(\Gamma)$ complements the support of $\mu_m(\theta)$. That is, λ should be chosen so that the tails of Γ are thicker than those of $\mu_m(\theta)$, which will allow us to detect deviations from compatibility, but not so large as to cause the statistic $\phi_m(\zeta)$ to have heavy tails. Indeed, if the tails of $\phi_m(\zeta)$ are too heavy, the implicit normality assumption made in BSL will be violated and can result in inefficient sampling.

Since there is no reason to believe a priori that θ and Γ are related, we take as our overall prior on ζ in R-BSL-M to be

$$\pi(\zeta) := \pi(\theta) \text{La}(0, \lambda).$$

As a default choice of prior, we suggest to select $\lambda = 0.5$, as this places most of the prior support for allowing up to ± 3 standard deviations shift in the mean for each summary statistic (see Figure 2).

3.2 Variance Compatibility and Prior Specification

While one approach to ensure compatibility is to adjust the mean of the simulated summaries, an alternative is to inflate the variance of the simulated summaries to ensure that $\eta(\mathbf{y})$ is always in the support of $\mu_m(\theta)$.⁹

Under regularity conditions and for fixed m , it is likely to be the case that the centered

⁹Recall that, if $\eta(\mathbf{y})$ is a sample average, the sample variance of $\mu_m(\theta)$ is converging to zero as either n or m diverges. Effectively, this has the effect of shrinking the support of $\mu_m(\theta)$.

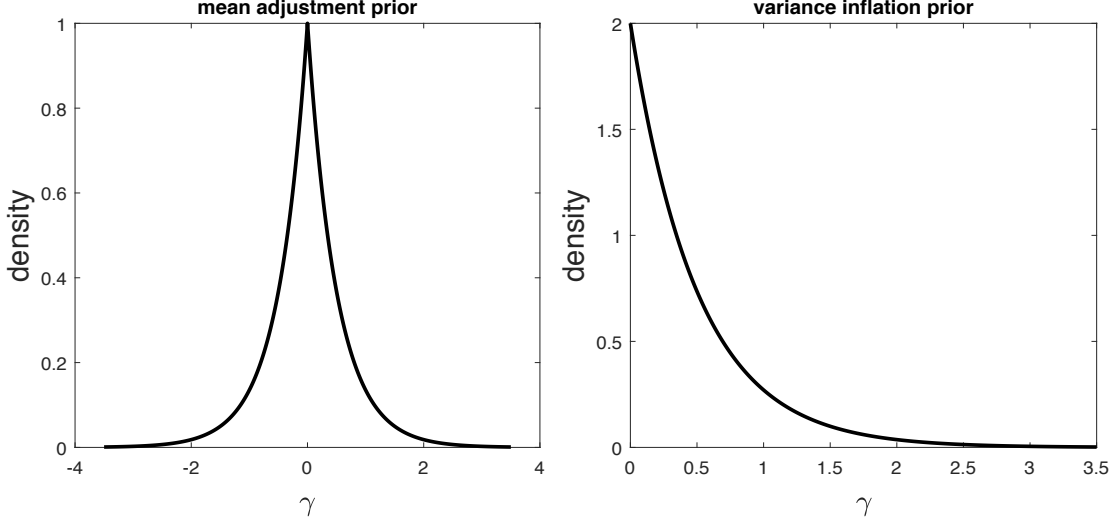


Figure 2: Default prior distribution on γ for mean adjustment (left) and variance inflation (right).

statistic $Z_{n,m}(\theta) := \{\mu_m(\theta) - b(\theta)\}$ behaves as $Z_{n,m}(\theta) = O_P(1/(m\sqrt{n}))$, with the variance of $Z_{n,m}(\theta)$ decreasing like $1/(mn)$. Consequently, for n (or m) large enough, if for a given value of θ the statistic $\{\eta(\mathbf{y}) - b(\theta)\}$ is more than a few standard deviations (as measured by $\Sigma_m^{1/2}(\theta)$) away from $\{\mu_m(\theta) - b(\theta)\}$, we can effectively view the summaries as being incompatible.

Given this characterization, an alternative approach to ensure that $\mu_m(\theta)$ is compatible with $\eta(\mathbf{y})$ is to artificially inflate the variance $\Sigma_m(\theta)$ so that the variance of $Z_{n,m}(\theta)$ never completely collapses to zero, and thus $\{\eta(\mathbf{y}) - b(\theta)\}$ can always be found in the support of $Z_{n,m}(\theta)$, albeit perhaps with small probability. More specifically, we propose to artificially inflate the variance used within BSL by adding to $\Sigma_m(\theta)$ the free parameters $\Gamma = (\gamma_1, \dots, \gamma_{d_\eta})^\top$.

Recalling $\zeta = (\theta^\top, \Gamma^\top)^\top$, a robust BSL procedure based on adjusting the variance can be implemented by re-defining the variance of the simulated statistics used within BSL to

be

$$V_m(\zeta) := \Sigma_m(\theta) + \begin{pmatrix} [\Sigma_m(\theta)]_{11}\gamma_1^2 & 0 & \dots & 0 \\ 0 & [\Sigma_m(\theta)]_{22}\gamma_2^2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & [\Sigma_m(\theta)]_{d_\eta d_\eta}\gamma_{d_\eta}^2 \end{pmatrix}, \quad (6)$$

where $[\Sigma_m(\theta)]_{ii}$ denotes the (i, i) element of $\Sigma_m(\theta)$. Given the structure of $V_m(\zeta)$, γ_i can be interpreted as an inflation factor operating on the standard deviations of the original BSL variance. An equivalent interpretation is that the i -th, $i = 1, \dots, d_\eta$, BSL variance is multiplied by the factor $1 + \gamma_i^2$. Using $V_m(\zeta)$ in place of $\Sigma_m(\theta)$ in the BSL posterior target, (1), and for $\pi(\zeta)$ an appropriate prior on ζ , the Robust BSL-variance (R-BSL-V) posterior is given as:

$$\hat{\pi}[\zeta|\eta(\mathbf{y})] \propto \bar{g}_n[\eta(\mathbf{y})|\zeta]\pi(\zeta),$$

where

$$\bar{g}_n[\eta(\mathbf{y})|\zeta] = \int \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta), V_m(\zeta)] \left\{ \prod_{i=1}^m g_n[\eta(\mathbf{z}^i) | \theta] \right\} d\eta(\mathbf{z}^1) \cdots d\eta(\mathbf{z}^m). \quad (7)$$

We note that the variance adjustment approach is operationally similar to using a tempered version of the synthetic likelihood, where the parameter Γ controls the “amount” of tempering, in that the value of Γ allow us to artificially fatten the tails of the likelihood. Tempered likelihoods in Bayesian inference are often suggested as a means of conducting robust inference in the context of model misspecification (Bissiri et al., 2016). While interesting, a thorough comparison between these two approaches is beyond the scope of this paper and is left for future research.

Prior Choice: Exponential Prior

Note that, by considering the standardization in (6), we ensure that each γ_i has the same scale and can be considered as unit-less. Moreover, similar to the case of the mean adjustment BSL approach, there is no reason to believe there is any *a priori* dependence between θ and Γ , so we can consider independent priors, i.e, $\pi(\zeta) := \pi(\theta)\pi(\Gamma)$. While several prior choices exist for Γ , following the arguments for the prior choice in the mean adjustment procedure, we need to choose a prior for the components of Γ so that there is a large amount of prior mass near the origin, and enough mass out in the tails to ensure we can detect incompatible summaries.

To this end, we consider independent exponential priors for each component γ_i , ($i = 1, \dots, d_\eta$), with common rate $\lambda > 0$:

$$\pi(\Gamma) := \prod_{i=1}^{d_\eta} \lambda e^{-\lambda \gamma_i} = \lambda^{d_\eta} e^{-\lambda \sum_{i=1}^{d_\eta} \gamma_i}.$$

The hyper-parameter λ should be chosen so that a large amount of prior mass is close to the origin, so as not to over-inflate the variance of the simulated summaries that are compatible.

While this choice of prior is not, strictly speaking, a shrinkage prior, it is still the case that we should observe some shrinkage like behavior for summaries that are compatible. That is, for the summaries that are compatible, this additional inflation by Γ is unnecessary and we expect that, for appropriate choices of λ , the addition of this component will not greatly affect the corresponding components in the variance. In contrast, for the summaries that are not compatible, this adjustment term is critical to ensure that the variance of the summaries is large enough to contain the observed summary $\eta(\mathbf{y})$. As a default choice of prior, we suggest to select a mean of $\lambda = 0.5$, as this places most of the prior support for allowing an additive inflation in the variance of up to 3 times the standard deviation of each summary statistic (see Figure 2).

3.3 Comparison of BSL and R-BSL

At this stage, it is useful to compare and contrast BSL and our robust approach to BSL under model misspecification to understanding why R-BSL will produce more reliable/robust statistical inferences under model misspecification. To this end, we now compare the behavior of the dominant terms within the BSL and R-BSL likelihoods under model misspecification, which will highlight the fundamental differences between the posteriors that can emerge in practice.

In the case of R-BSL-M, the behavior of the posterior is driven by the quadratic form

$$\|\phi_m(\zeta) - \eta(\mathbf{y})\|_{\Sigma_m(\theta)} = [\phi_m(\zeta) - \eta(\mathbf{y})]^\top \Sigma_m^{-1}(\theta) [\phi_m(\zeta) - \eta(\mathbf{y})].$$

The R-BSL-M posterior assigns higher probability mass to values of (θ, Γ) for which $\|\phi_m(\zeta) - \eta(\mathbf{y})\|_{\Sigma_m(\theta)}$ is “small”, i.e., values of (θ, Γ) so that $\phi_m(\zeta) = \mu_m(\theta) + \text{diag}[\Sigma_m^{1/2}(\theta)] \Gamma$ is close to $\eta(\mathbf{y})$.

The behavior of the R-BSL-V posterior is driven by the quadratic form

$$\|\mu_m(\theta) - \eta(\mathbf{y})\|_{V_m(\zeta)} = [\mu_m(\theta) - \eta(\mathbf{y})]^\top V_m^{-1}(\zeta) [\mu_m(\theta) - \eta(\mathbf{y})].$$

If there are many values of θ for which $\|\mu_m(\theta) - \eta(\mathbf{y})\|$ is already small, then variance inflation is not needed, and the resulting posterior for Γ will be uninformative (and resemble the prior); if there are no values of θ that make $\|\mu_m(\theta) - \eta(\mathbf{y})\|$ “small”, then $\|\mu_m(\theta) - \eta(\mathbf{y})\|_{V_m(\zeta)}$ can always be made small by choosing a large value of Γ . Consequently, when no value of θ exists for which $\mu_m(\theta)$ is close to $\eta(\mathbf{y})$, the R-BSL-V posterior will assign high posterior mass to values of θ for which $\mu_m(\theta)$ is as close as possible to $\eta(\mathbf{y})$, and values of Γ that ensure $\|\mu_m(\theta) - \eta(\mathbf{y})\|_{V_m(\zeta)}$ is small.

The posterior behavior of standard BSL is driven by the quadratic form

$$\|\mu_m(\theta) - \eta(\mathbf{y})\|_{\Sigma_m(\theta)} = [\mu_m(\theta) - \eta(\mathbf{y})]^\top \Sigma_m^{-1}(\theta) [\mu_m(\theta) - \eta(\mathbf{y})].$$

While R-BSL has in-built mechanisms to ensure that these quadratic forms can be made small, no such mechanism exists for BSL: by the inequality $\|a - b - c\| \geq \|b\| - \|a\| - \|c\|$,

$$\begin{aligned} \|\mu_m(\theta) - \eta(\mathbf{y})\|_{\Sigma_m(\theta)} &\geq \|b(\theta) - b_0\|_{\Sigma_m(\theta)} - \|b(\theta) - \mu_m(\theta)\|_{\Sigma_m(\theta)} - \|\eta(\mathbf{y}) - b_0\|_{\Sigma_m(\theta)} \\ &\geq \|b(\theta) - b_0\|_{\Sigma_m(\theta)} - o_p(1), \end{aligned}$$

where the second inequality comes from the fact that $b_0 := \text{plim}_n \eta(\mathbf{y})$ and $b(\theta) := \text{plim}_n \mu_m(\theta)$, for any $m \geq 1$. Under incompatibility, i.e., model misspecification, the term $\|b(\theta) - b_0\|_{\Sigma_m(\theta)}$ can be quite large, and is strictly positive in the limit, which results in a BSL posterior that is sensitive to the level/nature of model misspecification and which can ultimately be ill-behaved (for example, bi-modal); see Section 4.2 for a particular example. In contrast, since R-BSL ensures a form of compatibility can be achieved, the resulting posteriors will not be (particularly) sensitive to the level of model misspecification, and will be better behaved than their BSL counterpart.

3.4 Sampling Robust BSL

As demonstrated in Price et al. (2018), the standard BSL target posterior is given by

$$\hat{\pi}[\zeta \mid \eta(\mathbf{y})] \propto \bar{g}_n[\eta(\mathbf{y}) \mid \theta] \pi(\theta), \quad (8)$$

where

$$\bar{g}_n[\eta(\mathbf{y}) \mid \theta] = \int \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta), \Sigma_m(\theta)] \left\{ \prod_{i=1}^m g_n[\eta(\mathbf{z}^i) \mid \theta] \right\} d\eta(\mathbf{z}^1) \cdots d\eta(\mathbf{z}^m).$$

Price et al. (2018) use a Metropolis-Hastings algorithm to sample from (8) that proposes $\theta^* \sim q(\cdot \mid \theta)$ according to a Markov transition, and estimating $\bar{g}_n[\eta(\mathbf{y}) \mid \theta^*]$ unbiasedly through a single draw from $\prod_{i=1}^m g_n[\eta(\mathbf{z}^i) \mid \theta^*]$ and evaluating $\mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta), \Sigma_m(\theta)]$. Using pseudo-marginal MCMC arguments of Andrieu and Roberts (2009), substituting this estimator into a Metropolis-Hastings algorithm produces an algorithm that targets (8).

Our robust BSL methods operate on an extended state space over θ and Γ with target distribution

$\hat{\pi}[\zeta \mid \eta(\mathbf{y})] \propto \bar{g}_n[\eta(\mathbf{y}) \mid \zeta] \pi(\theta) \pi(\Gamma)$, where

$$\bar{g}_n[\eta(\mathbf{y}) \mid \zeta] = \int \mathcal{N}[\eta(\mathbf{y}); \phi_n(\zeta), \Sigma_n(\theta)] \left\{ \prod_{i=1}^m g_n[\eta(\mathbf{z}^i) \mid \theta] \right\} d\eta(\mathbf{z}^1) \cdots d\eta(\mathbf{z}^m) \quad (\text{Mean Adjustment}),$$

$$\bar{g}_n[\eta(\mathbf{y}) \mid \zeta] = \int \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta), V_m(\zeta)] \left\{ \prod_{i=1}^m g_n[\eta(\mathbf{z}^i) \mid \theta] \right\} d\eta(\mathbf{z}^1) \cdots d\eta(\mathbf{z}^m) \quad (\text{Variance Adjustment}).$$

To sample these target distributions we use a component-wise MCMC algorithm that updates, in turn, θ conditional on Γ and then Γ conditional on θ . The update for θ is the same as in standard BSL, but where the adjusted mean or inflated variance is computed as appropriate using the current value of Γ . As before, the update for θ involves generating m model simulations, $\eta(\mathbf{z}^i)$, $i = 1, \dots, m$.

The update for Γ holds the currently accepted model simulations fixed, and thus $\mu_m(\theta)$ and $\Sigma_m(\theta)$ are fixed within the update step for Γ . Each component of Γ , γ_j , for $j = 1, \dots, d_\eta$, is updated separately, conditional on the current values of the remaining components (denoted $\gamma_{/j}$). The full conditional distribution for γ_j is given by

$$\pi(\gamma_j^* \mid \theta, \mu_m(\theta), \Sigma_m(\theta), \gamma_{/j}) \propto \mathcal{N}[\eta(\mathbf{y}); \phi_m(\zeta^*), \Sigma_m(\theta)] \pi(\gamma_j^*) \quad (\text{Mean Adjustment})$$

$$\pi(\gamma_j^* \mid \theta, \mu_m(\theta), \Sigma_m(\theta), \gamma_{/j}) \propto \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta), V_m(\zeta^*)] \pi(\gamma_j^*) \quad (\text{Variance Inflation}),$$

where γ_j^* is a realisation of γ_j and $\zeta^* = [\theta, \gamma_j^*, \gamma_{/j}]$.

We sample this full conditional distribution using a slice sampler, and, in particular, the “stepping out” and “shrinkage” procedures detailed in Neal (2003). The appeal of the slice sampler is that the acceptance probability is one, and thus there are no tuning parameters that can affect the statistical efficiency. However, a stepping out width needs to be selected, which can impact the speed of the slice sampler. Given that the components of Γ are eventually scaled by the summary statistic standard deviation, and given that our prior choices effectively penalise large values, we expect each component of Γ to be

$O(1)$. Thus, we set the stepping out width to be 1, except for the lower bound of γ_j in the variance inflation, which is immediately set to 0; and hence the stepping out procedure is not required. We find this choice of width to be suitable, and since updating Γ does not require any model simulations, the slice sampler is very fast. Hence, importantly, our robust BSL methods do not require any additional tuning and the run-time per iteration for non-trivial applications is not noticeably slower. The full MCMC algorithm to sample from the R-BSL posteriors is provided in Algorithm 1.

It is important to note that under either R-BSL approach, we recover the original BSL target when

$$\gamma_1 = \cdots = \gamma_{d_\eta} = 0.$$

Therefore, if the model P_θ^n can generate summaries $\eta(\mathbf{z})$ that match $\eta(\mathbf{y})$, under an appropriate prior specification, the posterior $\hat{\pi}[\Gamma \mid \eta(\mathbf{y})]$ should not differ substantially from the prior, with most of the posterior mass located near the origin.

3.5 Theoretical Properties of R-BSL

If the model is compatible, what behavior should we expect from the R-BSL approach? Given the priors used in R-BSL, which place the majority of their mass near the origin, one would hope that when the compatibility condition (Definition 1) is satisfied, the introduction of the additional parameters Γ does not influence the BSL posterior for θ . We demonstrate that when Definition 1 is satisfied the asymptotic behavior of the R-BSL posterior of θ components behaves the same as the exact BSL posterior: namely, both posteriors concentrate all mass onto the value of θ that satisfies $b(\theta) = b_0$. In addition, we demonstrate that when Definition 1 is satisfied the R-BSL posteriors for the adjustment components converge to the prior. Consequently, we do not lose anything by using R-BSL when the model is correctly specified, but gain robustness to deviations from the modeling assumptions if the model is misspecified.

Before presenting the formal result, we must state some notation. Define P_0^n as the true distribution generating \mathbf{y} . The map $\eta : \mathbb{R}^n \rightarrow \mathbb{R}^{d_\eta}$, $d_\eta \geq d_\theta$, which defines the summary

Input : Summary statistic of the data, $\eta(\mathbf{y})$, the prior distribution, $\pi(\theta)$, the proposal distribution q , the number of iterations, T , and the initial value of the chain, θ^0, Γ^0 .

Output: MCMC sample $(\theta^0, \theta^1, \dots, \theta^T)$ and $(\Gamma^0, \Gamma^1, \dots, \Gamma^T)$ from the robust BSL posterior. Some samples can be discarded as burn-in if required.

- 1 Estimate $\mu_m(\theta^0)$ and $\Sigma_m(\theta^0)$ via m independent model simulations at θ^0
- 2 Compute $\phi_m(\zeta^0) = \mu_m(\theta^0) + \text{diag} \left[\Sigma_m^{1/2}(\theta) \right] \Gamma^0$ (mean adjustment) or $V_m(\zeta^0)$ (variance inflation) defined in (6).
- 3 Compute robust synthetic likelihood $L^0 = \mathcal{N}[\eta(\mathbf{y}); \phi_m(\zeta^0), \Sigma_m(\theta^0)]$ (mean adjustment) or $L^0 = \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta^0), V_m(\zeta^0)]$ (variance inflation)
- 4 **for** $i = 1$ **to** T **do**
- 5 %%%% Obtain Γ^i via the following, which does not require any model simulations
- 6 **for** $j = 1$ **to** d_η **do**
- 7 Update γ_j^i with a slice sampler with target density

$$\pi(\gamma_j^i | \theta^{i-1}, \mu_m(\theta^{i-1}), \Sigma_m(\theta^{i-1}), \gamma_1^i, \dots, \gamma_{j-1}^i, \gamma_{j+1}^{i-1}, \dots, \gamma_{d_\eta}^{i-1}) \propto$$

$$\mathcal{N}[\eta(\mathbf{y}); \phi_m(\zeta^*), \Sigma_m(\theta^{i-1})] \pi(\gamma_j^i) \text{ (mean adjustment) or}$$


$$\mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta^{i-1}), V_m(\zeta^*)] \pi(\gamma_j^i) \text{ (variance inflation) where}$$

$$\zeta^* = [\theta^{i-1}, \gamma_1^i, \dots, \gamma_{j-1}^i, \gamma_j^i, \gamma_{j+1}^{i-1}, \dots, \gamma_{d_\eta}^{i-1}]$$
- 8 **end**
- 9 Denote $L^i = \mathcal{N}[\eta(\mathbf{y}); \phi_m([\theta^{i-1}, \Gamma^i]), \Sigma_m(\theta^{i-1})]$ (mean adjustment) or
 $L^i = \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta^{i-1}), V_m([\theta^{i-1}, \Gamma^i])]$ (variance inflation)
- 10 %%%% Update θ^i conditional on Γ^i
- 11 Draw $\theta^* \sim q(\cdot | \theta^{i-1})$
- 12 Estimate $\mu_m(\theta^*)$ and $\Sigma_m(\theta^*)$ via m independent model simulations at θ^*
- 13 Compute $\phi_m(\zeta^*)$ (mean adjustment) or $V_m(\zeta^*)$ (variance inflation) defined in (6),
where $\zeta^* = [\theta^*, \Gamma^i]$
- 14 Compute proposed adjusted synthetic likelihood $L^* = \mathcal{N}[\eta(\mathbf{y}); \phi_m(\zeta^*), \Sigma_m(\theta^*)]$
(mean adjustment) or $L^* = \mathcal{N}[\eta(\mathbf{y}); \mu_m(\theta^*), V_m(\zeta^*)]$ (variance inflation)
- 15 Compute Metropolis-Hastings ratio:

$$r = \frac{L^* \pi(\theta^*) q(\theta^{i-1} | \theta^*)}{L^i \pi(\theta^{i-1}) q(\theta^* | \theta^{i-1})}$$
- 16 **if** $\mathcal{U}(0, 1) < r$ **then**
- 17 Set $\theta^i = \theta^*$, $\mu_m(\theta^i) = \mu_m(\theta^*)$ and $\Sigma_m(\theta^i) = \Sigma_m(\theta^*)$
- 18 **else**
- 19 Set $\theta^i = \theta^{i-1}$, $\mu_m(\theta^i) = \mu_m(\theta^{i-1})$ and $\Sigma_m(\theta^i) = \Sigma_m(\theta^{i-1})$
- 20 **end**
- 21 **end**

Algorithm 1: Robust MCMC BSL.

statistics used in the procedure, satisfies $\eta(\mathbf{y}) \sim G_n^0$, where G_n^0 denotes the projection of P_n^0 under the map η , and denote by g_n^0 the density of G_n^0 . Likewise, recall that $G_n[\cdot|\theta]$ denotes the projection of P_n^θ under the map η . For real-valued sequences $\{a_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$: $a_n \lesssim b_n$ denotes $a_n \leq Cb_n$ for some finite $C > 0$ and all n large, $a_n \gtrsim b_n$ denotes $a_n \geq Cb_n$ for some finite $C > 0$, and $a_n \asymp b_n$ implies $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. The terms O_P and o_P have their usual connotations and the notation \Rightarrow denotes weak convergence in distribution.

Recall that, $b(\theta) = \mathbb{E}[\eta(\mathbf{y})|\theta]$ and $\Sigma(\theta) := \mathbb{E}[\{\eta(\mathbf{y}) - b(\theta)\} \{\eta(\mathbf{y}) - b(\theta)\}' | \theta]$. We impose the following regularity conditions. 

Assumption 1. *There exists a sequence of positive real numbers v_n diverging to ∞ such that, for some distribution Q on \mathbb{R}^{d_η} and some vector $b_0 \in \mathbb{R}^{d_\eta}$, $v_n [\eta(\mathbf{y}) - b_0] \Rightarrow Q$, under P_n^0 .*

Assumption 2. *(i) The sequence $\{v_n\}_{n \geq 1}$ is such that, for all $\theta \in \Theta$ and some n large enough, there exists constants c_1, c_2 , $c_1 \leq c_2$, satisfying: $0 < c_1 \leq \|v_n \Sigma_n(\theta)\|_* \leq c_2 < \infty$, for some matrix norm $\|\cdot\|_*$; (ii) For all $\theta \in \Theta$ and some n large enough, the $d \times d$ -matrix $A_n(\theta)$ is continuous in θ .*

Assumption 3. *There exists a deterministic map $\theta \mapsto b(\theta)$, such that, for all $\theta \in \Theta$, and for constants $\alpha, u_0 > 0$, for all $0 < u < u_0 v_n$, $G_n[\|v_n \{\eta(\mathbf{y}) - b(\theta)\}\| > u | \theta] \leq c(\theta) u^{-\alpha}$, uniformly for $n \geq 1$ and where $\int_\Theta c(\theta) \pi(\theta) d\theta = O(1)$.*

Assumption 4. *(i) There exists some $\tau > 0$ such that, for all $0 < u < u_0 v_n$, the prior probability satisfies $\Pi[\|b(\theta) - b_0\| \leq u] \asymp u^\tau$. (ii) The prior density $\pi(\theta)$ is continuous and satisfies $\pi(\theta_0) > 0$.*

Assumption 5. *(i) The map $\theta \mapsto b(\theta)$ is continuous and injective, with $b(\theta_0) = b_0$ for some $\theta_0 \in \Theta$, and satisfies: $\|\theta - \theta_0\| \leq L \|b(\theta) - b_0\|^\kappa$ on some open neighbourhood of θ_0 with $L > 0$ and $\kappa > 0$.*

Assumption 6. *If Assumption 11 is satisfied, for any $\epsilon > 0$, there exists $u, \delta > 0$ and a set V_n such that, for all $\theta \in \{\theta : \|b(\theta) - b_0\| \leq uv_n^{-1}\}$, $V_n \subset \{\eta \in \mathbb{R}^{d_\eta} : g_n^0(\eta) \lesssim g_n(\eta | \theta)\}$ where $P_n^0(V_n^c) < \epsilon$.*

The above assumptions are similar to those used in Marin et al. (2014) to deduce the behavior of Bayes factors in situations where inference is conditioned on summary statistics, as opposed to the entire data set. Due to space constraints, we defer a detailed discussion on these assumptions to Section 2 of the supplementary material.

The following result, the proof of which is also given in the supplemental material, describes the theoretical behavior of the R-BSL posterior under the above assumptions.

Proposition 1. *Under Assumption 7-12, for any $\delta > 0$,*

$$\Pi [\|\theta - \theta_0\| \leq \delta | \eta(\mathbf{y})] = 1 + o_P(1).$$

Moreover, for any $A \subseteq \mathcal{G}$:

$$\Pi [\Gamma \in A | \eta(\mathbf{y})] = \Pi[\Gamma \in A] + o_P(1).$$

Proposition 2 demonstrates that under compatibility the posterior for the model parameters, θ , are asymptotically unaffected by the introduction of the Γ components: the R-BSL posterior for θ concentrates all posterior mass onto the value θ_0 , the value under which $b(\theta) = b_0$. Consequently, from Proposition 1 in Frazier et al. (2019), the R-BSL and BSL posteriors behavior similarly when the model is compatible (i.e., when Definition 1 is satisfied) and implies that R-BSL does not pay a penalty for being robust to model misspecification if the model is correctly specified.¹⁰

Proposition 2 also demonstrates that, under compatibility, the posterior for the Γ components converge to the prior. This implies that under correct model specification the resulting posteriors for the Γ components will not asymptotically concentrate on the origin. As discussed in Bhattacharya et al. (2012), this is not surprising given the relatively mild shrinkage priors placed on the adjustment components. We conjecture that if stronger shrinkage priors were employed, posterior concentration, toward the origin, for these com-

¹⁰We note that determining the theoretical behavior of BSL and R-BSL when compatibility is not in evidence is a significant undertaking, and a novel research topic in its own right. Therefore, we leave such technical details for future research.

ponents could also be achieved. However, the use of these more complex priors could create issues within the sampling.

When the compatibility condition in Definition 1 is not satisfied, the R-BSL posterior for Γ will deviate from the prior. Therefore, the result of Proposition 2 can be used to determine the level of model misspecification by comparing the difference between the R-BSL posterior for Γ and the prior for Γ . While visual detection will often be enough to determine if any meaningful differences between these two exist, any norm on the space of probability measures could be used to quantify this discrepancy.

By analyzing the posterior elements of Γ that differ from the prior, we can deduce precisely which of the summary statistics the assumed model can not match, i.e., which summaries are not compatible. This information can then be incorporated into subsequent modelling steps to construct a model that can more adequately capture the observed data. In this way, R-BSL can also be used as a model criticism device to help researchers locate discrepancies between the assumed model and the observed data.

4 Examples

In this section, we first consider two toy examples that demonstrate the statistical benefits of R-BSL, over BSL. Next, we apply BSL and R-BSL to conduct inference in a challenging problem in ecology: using real data to model invasive toad populations. Collectively, these examples echo the analysis in Section 3.3 and demonstrate that R-BSL delivers more robust statistical inferences than BSL under model misspecification. In all examples we use the default priors shown in Figure 2.

4.1 Toy Example 1 Continued: Contaminated Normal Model

In this section, we analyze the performance of R-BSL in the contaminated normal example. We refer the reader to Section 2.3 for a detailed description of the underlying Monte Carlo design. Following the analysis in Section 2.3, we use precisely the same “observed” data

in Section 2.3, and apply R-BSL to this data.¹¹ The acceptance rates for BSL, and R-BSL are plotted in Panel A of Figure 3, while the resulting posterior median and credible sets are given in Panels B, C and D.

The results in Panels B, C and D of Figure 3 demonstrate that R-BSL yields reliable statistical inferences on θ regardless of the level of model misspecification. In stark contrast to BSL, across all the experiments, the posterior means and credible sets for R-BSL are virtually unaffected by the data contamination. Panel A of Figure 3 demonstrates that the acceptance rate for the variance adjustment R-BSL approach is only slightly affected by the level of model misspecification, while the mean adjustment version does display some degradation but still maintains acceptance rates above 5% in all cases. In comparison, when the sample standard deviation is equal to 2, the BSL acceptance rate is less than 0.01%.

In addition, we note that numerically implementing R-BSL is not much more computationally costly than implementing BSL. In this example, the execution time required to sample the R-BSL posterior (via Algorithm 1) was only 40% slower than the execution time required to sample the BSL posterior, even though R-BSL is conducting inference on three times as many parameters. Moreover, the additional time per MCMC iteration for R-BSL is even less noticeable in realistic examples where model simulation is non-trivial.

In the supplementary material, we analyze the adjustment components for this example and compare the repeated sampling behavior of BSL and R-BSL across different levels of model misspecification. As expected, the R-BSL posteriors for the adjustment components demonstrate that the model is unable to match the second summary statistic, while the posteriors associated with the first adjustment component are indistinguishable from the prior. See Figure 1 in the supplementary appendix for full details. The repeated sampling results demonstrate that R-BSL behaves similarly to BSL when the model is correctly specified, but yields more accurate estimators when the model is misspecified. See Table 1 in the supplementary appendix for full details.

¹¹We again use a RWMH algorithm that is initialized at $\theta = 0$, and we run the sampler for 25,000 iterations, with the first 10,000 discarded for burn-in.

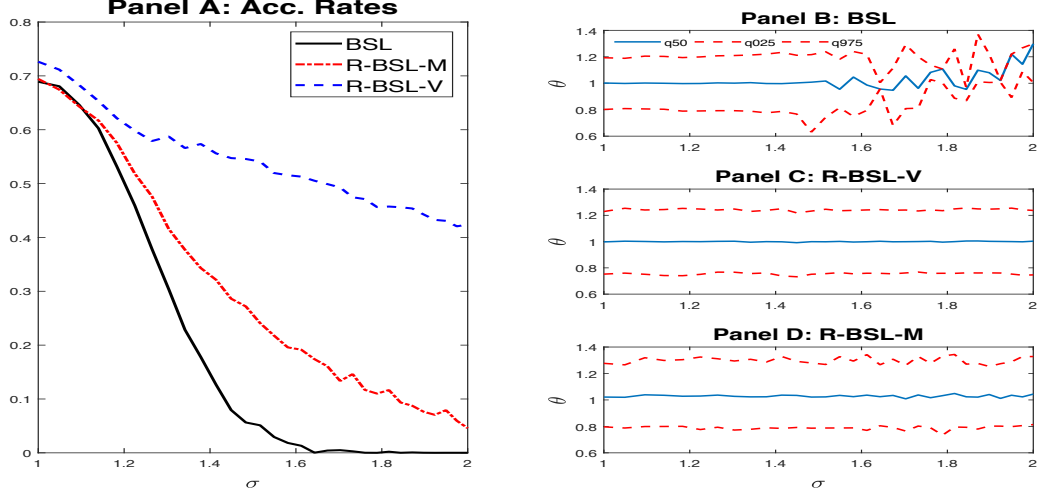


Figure 3: Panel A gives the acceptance rates for the different BSL procedures across the different data sets. Panels B, C and D give the posterior median and 95% credible sets for R-BSL and compares these with those obtained from BSL. We recall that the value of σ denotes the value of the sample (and population) standard deviation.

4.2 Toy Example 2: Moving Average Model

A common toy example used to demonstrate approximate inference methodology is the moving average (MA) model. The researcher believes \mathbf{y} is generated according to an MA(1) model:

$$z_t = e_t + \theta e_{t-1}, \quad (9)$$

and the unknown parameter θ satisfies $|\theta| < 1$, while our prior information on θ uniform over $(-1, 1)$. A useful choice of summary statistics for the MA(1) model are the sample autocovariances $\eta_j(\mathbf{z}) = \frac{1}{T} \sum_{t=1+j}^T z_t z_{t-j}$, for $j \in \{0, 1, 2\}$. Let $\eta(\mathbf{z})$ denote the summaries $\eta(\mathbf{z}) = (\eta_0(\mathbf{z}), \eta_1(\mathbf{z}), \eta_2(\mathbf{z}))^\top$. Under the DGP in equation (9), it can be shown that the summaries $\eta(\mathbf{z})$ satisfy

$$\eta(\mathbf{z}) \xrightarrow{P} b(\theta) := \begin{pmatrix} 1 + \theta_1^2 & \theta_1 & 0 \end{pmatrix}^\top.$$

While the researcher believes the data is generated according to an MA(1) model, the

actual DGP for \mathbf{y} evolves according to the stochastic volatility (SV) model

$$y_t = \exp(h_t/2)u_t, \quad h_t = \omega + \rho h_{t-1} + v_t \sigma_v, \quad (10)$$

where $0 < \rho < 1$, $0 < \sigma_v < 1$, u_t and v_t and both iid standard normal. In this case, if one takes $\eta(\mathbf{y}) = (\eta_0(\mathbf{y}), \eta_1(\mathbf{y}), \eta_2(\mathbf{y}))^\top$, under the DGP in (10),

$$\eta(\mathbf{y}) \xrightarrow{P} b_0 := \left(\exp\left(\frac{\omega}{1-\rho} + \frac{1}{2} \frac{\sigma_v^2}{1-\rho^2}\right), \quad 0, \quad 0 \right)^\top.$$

For any value of ω, σ_v and ρ such that $\exp\{\omega/(1-\rho) + \frac{1}{2}\sigma_v^2/(1-\rho^2)\} \neq 1$, the model is not compatible (and hence misspecified in the BSL sense); i.e., for any value of θ , $\|b(\theta) - b_0\| > 0$. From the definition of $b(\theta)$ and b_0 , it also follows that the value that minimizes $\|b(\theta) - b_0\|$ is $\theta = 0$, and it is this value onto which we would expect the R-BSL posterior to concentrate.

To understand how BSL and R-BSL perform in this misspecified model, we enact the following Monte Carlo experiment: we generate $n = 100$ observations from the SV model in (10) with parameter values $\omega = -0.76$, $\rho = 0.90$ and $\sigma_v = 0.36$, and use BSL, and R-BSL to conduct inference on θ in the misspecified MA(1) model. The R-BSL approach uses $m = 50$ simulated data sets to estimate the mean and the variance, while BSL uses $10 \times m = 500$ simulated data sets.

Under the true DGP in (10), the first two auto-correlations are zero for all values of $(\omega, \rho, \sigma_v)^\top$. Hence, we would expect that R-BSL will detect incompatibility in the first summary statistic, the sample variance, while the corresponding adjustment components for the other summaries would be indistinguishable from the prior, given in Figure 2.

For R-BSL we consider starting values obtained from the maximum likelihood estimators of the MA(1) model, and posterior draws for θ are obtained via a random-walk Metropolis sampler with fixed variance of 0.1. We run the MCMC sampler for 100,000 iterations and discard the first 10,000 for burn-in.

The nature of the model misspecification in this example results in a standard BSL

posterior that is ill-behaved; i.e., it is bi-modal with well separated modes, and the MCMC struggles to move between modes. Thus, given that there is only a single parameter, we use importance sampling based on 100,000 samples from the prior for θ to obtain the BSL posterior. The effective sample sizes of the importance sampling approximations over the 50 datasets is roughly between 50–500, allowing us to reasonably estimate the standard BSL posterior.

Under this Monte Carlo design, we generate fifty replications from the DGP in (10) and apply BSL and R-BSL to the ‘observed data’. The average acceptance rates across the replicated data sets for R-BSL-V and R-BSL-M are 38% and 26%, respectively. We display the posteriors for θ from each method, and across each data set, in Panels A-C of Figure 4.¹² Both R-BSL-V and R-BSL-M display significant posterior concentration around $\theta = 0$, while the BSL posterior is bi-modal and has little posterior mass around $\theta = 0$. The differences between the posteriors highlights the results discussed in Section 3.3, where we argued that the BSL and R-BSL posteriors can be dissimilar when the model is misspecified.

We now examine the marginal posteriors for the different adjustment components across R-BSL-V and R-BSL-M, which are given in Figure 5, across the fifty replications. The top row of Figure 5 gives the posterior densities of γ_1 , γ_2 and γ_3 for R-BSL-M and the bottom row corresponds to the same components for R-BSL-V. The results demonstrate that, as suggested, the model can not reliably match the first summary statistic, the sample variance, while the posteriors corresponding to the first and second-order autocorrelations do not significantly differ from their priors across all replicated data sets.

The accuracy of the BSL and R-BSL point estimators, across the repeated samples, is analyzed in Table 1. For each data set we run BSL, R-BSL-M, and R-BSL-V and calculate the bias (Bias), root mean squared error (RMSE), posterior credible set length (Len), and

¹²To deal with excess autocorrelation in the MCMC chains, the results in Panel A and B of Figure 4 have been thinned by taking every 100th sample. This significantly reduced the autocorrelation in the chain and permits smoother density estimates. Furthermore, we note here that the MCMC chains associated with each of the, non-thinned, posteriors satisfies the convergence diagnostic proposed in Geweke (1992) at any reasonable level of significance.

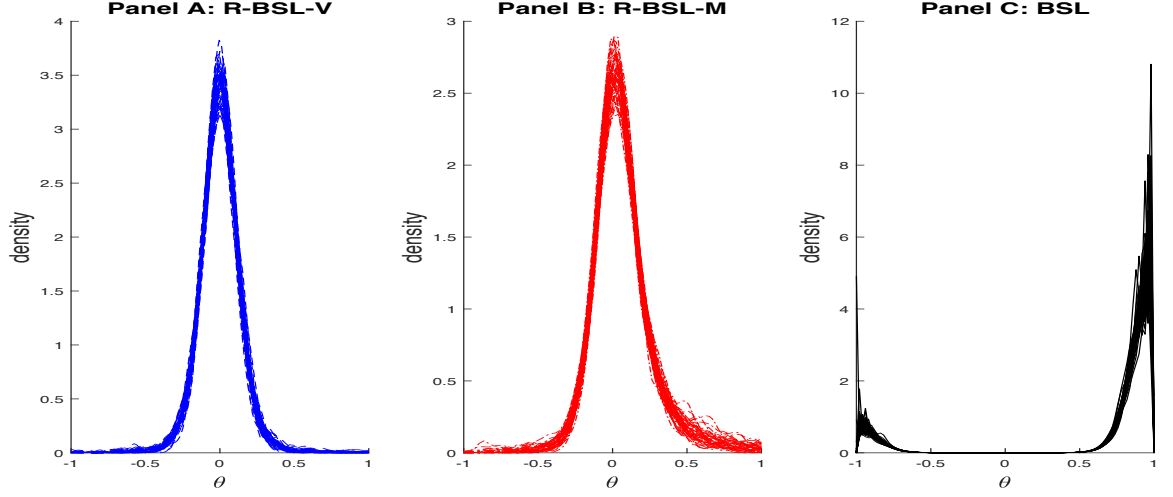


Figure 4: Posteriors for BSL, R-BSL-M and R-BSL-V for θ in the misspecified MA(1) model across fifty replicated data sets.

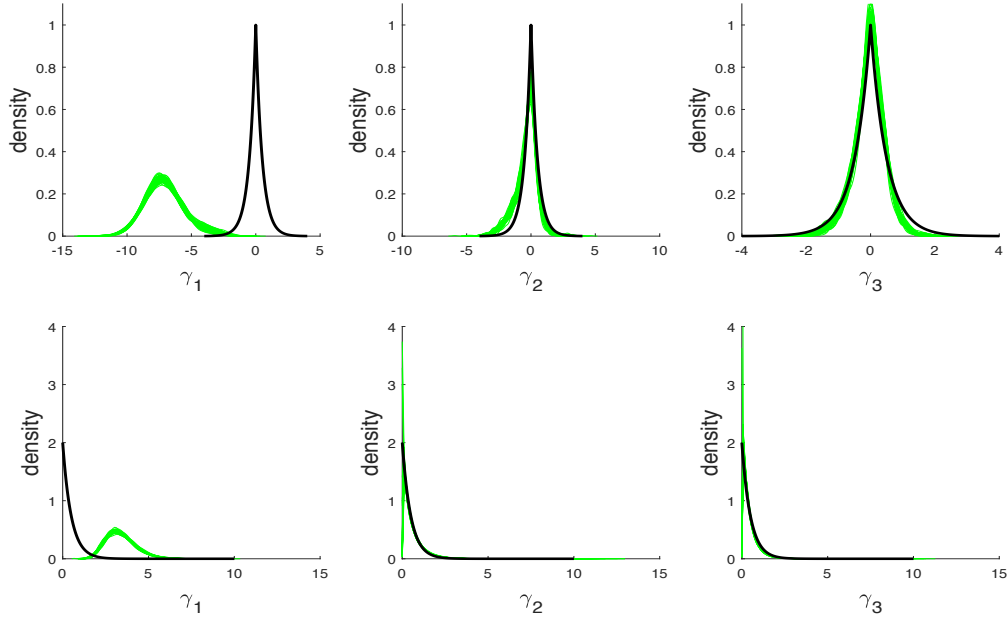


Figure 5: The top panels plot the R-BSL-M posteriors for $\gamma_1, \gamma_2, \gamma_3$, thin lines, and the corresponding priors, thick lines, across the fifty replicated data sets. The bottom panels displays the same information for R-BSL-V.

Monte Carlo coverage (COV), all relative to the pseudo-true value $\theta = 0$. The results are displayed in Table 1, and demonstrate that R-BSL-V yields the most accurate point

estimators, as measured by both bias and RMSE, followed by R-BSL-M. The bi-modal nature of the BSL posterior leads to a significantly biased point estimator, and ensures that the resulting credible set length and Monte Carlo coverage are not entirely meaningful. Hence, we do not report these quantities for BSL in Table 1.

The results in Figure 4 and Table 1 unequivocally demonstrate that when the model is misspecified R-BSL yields more reliable statistical inferences than those obtained by BSL.

Table 1: Summary measures for posterior mean accuracy, calculated as averages across the replications. RMSE- root mean squared error, BIAS- bias across the replications, LEN- credible set length, COV- Monte Carlo coverage. Due to the bi-modal BSL posterior, we do not report the resulting Monte Carlo coverage (COV) or credible set length (LEN) for BSL.

	BSL	R-BSL-V	R-BSL-M
RMSE	0.306	0.006	0.076
BIAS	0.305	-0.001	0.073
LEN	N/A	0.544	0.979
COV	N/A	100%	100%

4.3 Toad Example

4.3.1 Background

We consider an individual-based model of a species called Fowler’s Toads (*Anaxyrus fowleri*) developed by Marchand et al. (2017), which was also analysed by An et al. (2020). Here we give very brief details, with more information in Marchand et al. (2017) and An et al. (2020).

The model assumes that a toad hides in its refuge site in the daytime and moves to a randomly chosen foraging place at night. GPS location data are collected on n_t toads for n_d days, i.e. the observation matrix \mathbf{Y} is of dimension $n_d \times n_t$ ($n_t = 66$ and $n_d = 63$ here). Then \mathbf{Y} is summarised to 4 sets comprising the relative moving distances for time lags of 1, 2, 4, 8 days. For instance, \mathbf{y}_1 consists of the displacement information of lag 1 day, $\mathbf{y}_1 = \{|\mathbf{Y}_{i,j} - \mathbf{Y}_{i+1,j}|; 1 \leq i \leq n_d - 1, 1 \leq j \leq n_t\}$.

Simulating from the model involves two distinct processes. For each toad, we first generate an overnight displacement, Δy , then mimic the returning behaviour with a sim-

plified model. The overnight displacement is assumed to belong to the Lévy-alpha stable distribution family, with stability parameter α and scale parameter δ . The total returning probability is a constant p_0 , if a return occurs on day i , $1 \leq i \leq m$, then the return site is the same as the refuge site on day i , where i is selected randomly from $1, 2, \dots, m$ with equal probability. Here we consider both simulated and real datasets. For the synthetically generated data we take $\theta = (\alpha, \delta, p_0)^\top = (1.8, 45, 0.6)^\top$, which is informed by the parameter estimates obtained in Marchand et al. (2017). We use a uniform prior over $(1, 2) \times (0, 100) \times (0, 0.9)$. Marchand et al. (2017) consider three variations on the model. Here, we consider their ‘Model 2’ since there is a strong indication from their results that this model is not able to recover some of the chosen summary statistics.

As in Marchand et al. (2017), the dataset of displacements is split into two components. If the absolute value of the displacement is less than 10 metres, it is assumed the toad has returned to its starting location. For the summary statistic, we consider the number of toads that returned (Marchand et al. (2017)). For the non-returns (absolute displacement greater than 10 metres) we consider a larger collection of summaries. We calculate the log difference between adjacent p -quantiles with $p = 0, 0.1, \dots, 1$ and also the median. These statistics are computed separately for the four time lags. This results in 48 statistics in total, which is hard to handle for conventional ABC methods. An et al. (2020) demonstrate that BSL is computationally efficient enough to analyse simulated data for this application with a similar number of summary statistics.

For the γ parameters of the mean adjustment and variation inflation procedures we use the same priors as the previous example.

4.3.2 Results

We first consider the simulated dataset, where we use $n = 300$ simulations to estimate the synthetic likelihood at each MCMC iteration. Standard BSL, together with the two incompatibility extensions, produce approximate posteriors shown in Figure 6. As can be seen, the adjustments produce posteriors remarkably similar to BSL with slightly inflated

variances. The MCMC acceptance rates for BSL, R-BSL-M and R-BSL-V are 11%, 9% and 22%, respectively. As consistent with previous results, the variance adjustment improves the computational efficiency, even when the model is correctly specified. The posterior distributions for γ of R-BSL are shown in Figure 7. In all cases, the posteriors are not too dissimilar to the prior.

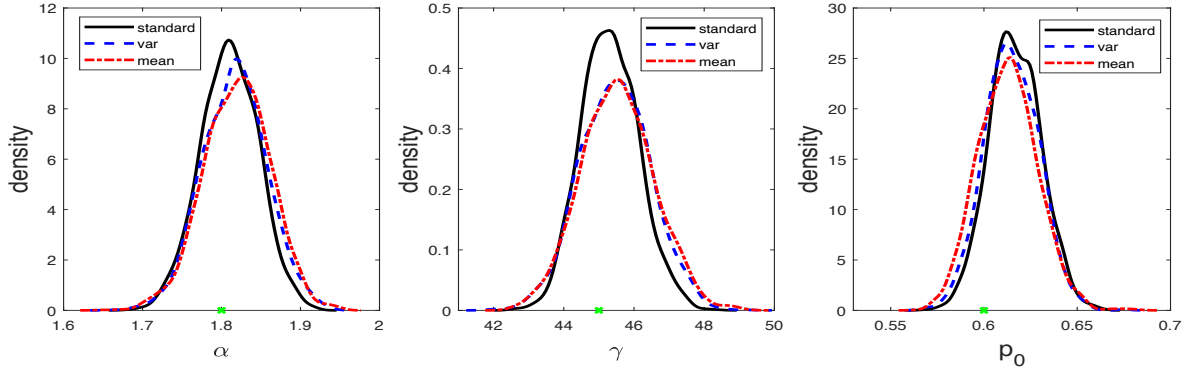


Figure 6: Univariate posterior distributions for the parameters when applying BSL (solid), R-BSL-V (dash) and R-BSL-M (dot-dash) to simulated data for the toad example. True parameter values are shown as crosses.

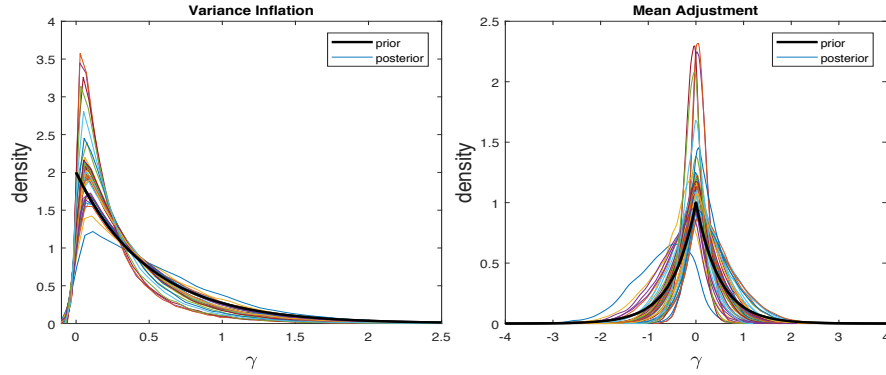


Figure 7: Posterior distributions for γ for R-BSL-M (right) and R-BSL-V (left) applied to the simulated data for the toad example. The thick line is the prior and the thin lines are the posteriors for the components of γ .

For the real data, we required $n = 2000$ simulations for estimating the synthetic likelihood to obtain an acceptance rate of 9% for BSL. However, the chain still suffered from periods of stickiness. In contrast, with only $n = 500$, the R-BSL-M and R-BSL-V produce acceptance rates of 7% and 15%, respectively, without substantial stickiness. The variance

inflation method offers a computational improvement of about one order of magnitude over BSL when accounting for both acceptance rate and number of simulations.

The posterior distributions for the components of γ for the R-BSL methods are shown in Figure 8. It is evident from the plots that our methods have identified that there are three or four statistics that the model is not compatible with. The statistic with the largest incompatibility is the number of returns for lag 1. For R-BSL-V, the 95% posterior predictive interval for this statistic is (262, 346) with an observed value of 234. Other statistics showing some incompatibility are the first quantile differences of the non-returns for lags 3 and 4. Figure 9 confirms that the observed data are not consistent with the posterior predictive distribution of the (log) non-return distances for lags 3 and 4, in that the model generally predicts larger non-return distances. The mean adjustment results are similar (not shown). Our adjustment methods permit in-depth analyses such as these and may provide practitioners valuable information for improving the model.

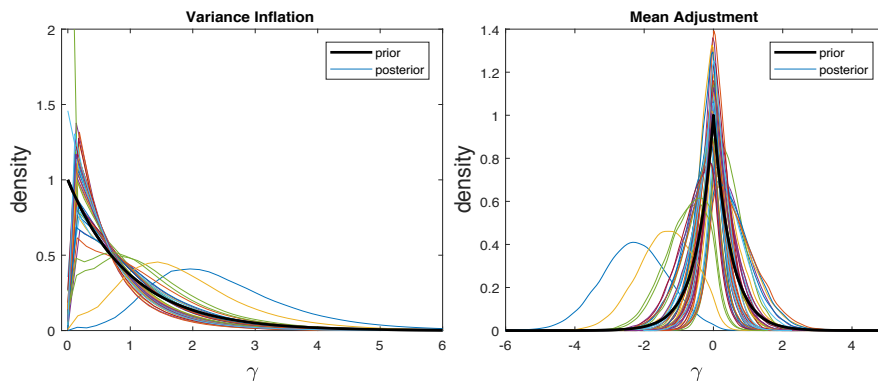


Figure 8: Posterior distributions for γ for R-BSL-M (right) and R-BSL-V (left) applied to the real data for the toad example. The thick line is the prior and the thin lines are the posteriors for the components of γ .

The bivariate posterior distributions for the parameters based on the adjustment methods are shown in Figure 10. It is evident that the estimated posterior distributions are similar, with the R-BSL-M posteriors slightly more concentrated than R-BSL-V. Univariate posteriors for all BSL approaches are shown in Figure 11. Comparing the univariate posteriors, we see that there is substantive disagreement between BSL and R-BSL. The R-BSL posteriors generally have fatter tails than the BSL posteriors, which indicates the

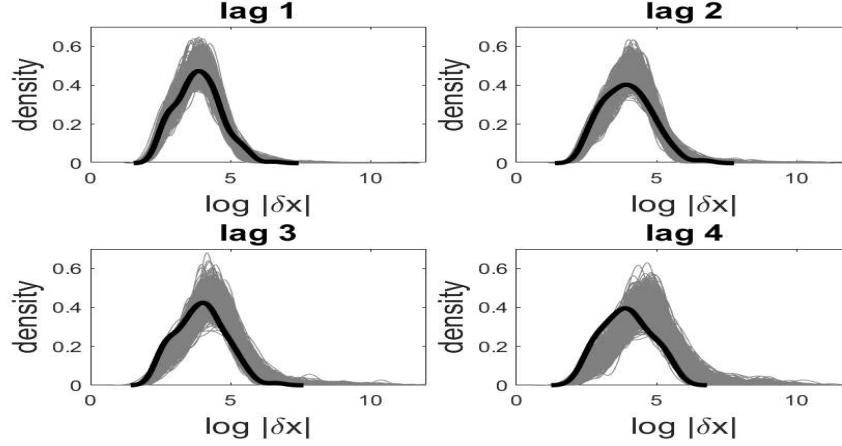


Figure 9: Posterior predictive distributions of the log non-returns for the four lags based on R-BSL-V. The thick line is the distribution for the observed data.

presence of model misspecification, and are centered over different regions of the support (especially for α and p_0). This empirical evidence reinforces the analysis in Section 3.3, where we argued that the R-BSL and BSL posteriors need not agree under model misspecification.

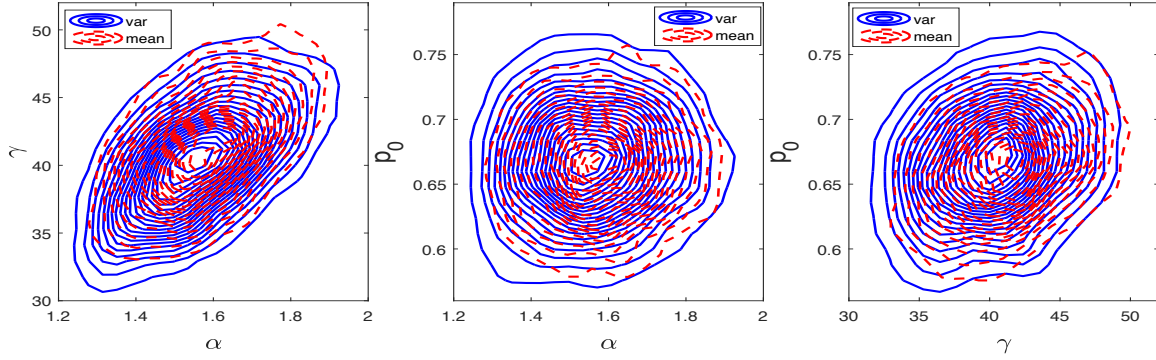


Figure 10: Bivariate posterior distributions visualised as contour plots for the parameters based on R-BSL-M (right) and R-BSL-V (left) applied to the real data for the toad example.

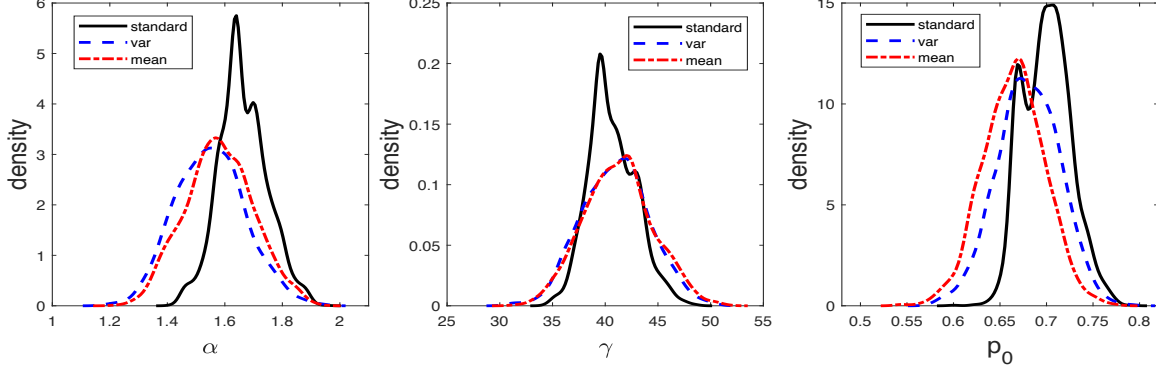


Figure 11: Univariate posterior distributions for the parameters when applying BSL (solid), R-BSL-V (dash) and R-BSL-M (dot-dash) to real data for the toad example.

5 Discussion

This paper has made two significant contributions to the literature on approximate Bayesian methods. Firstly, to our knowledge, this is the first piece of research to demonstrate that, similar to approximate Bayesian computation (ABC), Bayesian synthetic likelihood (BSL) can deliver unreliable inference when the assumed model is misspecified. Secondly, to circumvent the poor behavior of BSL in these settings, we have proposed a modification of BSL that displays robustness to model misspecification. Several Monte Carlo and empirical examples are used to illustrate the performance of this new method, with the results demonstrating both the statistical and computational benefits of this new approach when the model is misspecified.

In addition to delivering more accurate statistical inference under model misspecification, this new approach also allows the user to detect precisely which summary statistics are incompatible with the assumed data generating process. Incorporating this information within subsequent rounds of model building could lead to better models that can more accurately capture the behavior exhibited by the observed summary statistics. In this sense, the robust BSL approach can be viewed as a BSL version of the model criticism approach of Ratmann et al. (2009). In the context of ABC, Ratmann et al. (2009) propose an approach to detect aspects of the model that the summary statistics can not adequately capture. Their approach relies on treating the ABC tolerance as an unknown parame-

ter, and augmenting the original ABC inference problem with this additional parameter. The authors argue that posterior realizations for the tolerance parameter that are “large” indicate the possibility of a mismatch between the model and the observed data.

While useful, in the case of multivariate summaries the approach of Ratmann et al. (2009) requires a tolerance parameter for each summary statistic used in the analysis, with posterior inference then required on the full set of model parameters and tolerance parameters. Therefore, even for a moderate number of summaries, this approach can exacerbate the underlying curse-of-dimensionality in ABC, as it pertains to both the dimension of summaries and the number of parameters in the analysis. For instance, the empirical example in Section 4.3 employed 48 summary statistics, which is much larger than can reliably be considered by standard implementations of ABC. In addition, the approach considered herein has as a direct benchmark with which to gauge the impact of misspecification on the summaries, namely the prior distribution of the adjustment components. If the corresponding posterior for the adjustment component in the robust version of BSL does not resemble the prior, this is strong evidence that this summary can not be matched by the assumed model. If one wished to put a numerical value, or conduct a formal hypothesis test, on the difference between the prior and posterior, any number of techniques could be used.

The examples illustrate that, in particular, the variance inflation approach can significantly improve the MCMC acceptance rate, under model misspecification (relative to standard BSL). The variance inflation approach bears some resemblance to MCMC ABC approaches that assign a distribution to the ABC tolerance to facilitate MCMC mixing by proposing a relatively large tolerance value (e.g., Bortot et al. 2007). However, improving mixing is not our primary focus, but is simply a useful by-product. Our R-BSL approaches may also be useful for initial explorations of the parameter space when it is not known where the bulk of the posterior support is since, even for a correctly specified model, a poor parameter value will not be able to recover the observed statistic.

As with BSL, our R-BLS approach requires that a Gaussian approximation to the

summaries is reasonable. In cases where it is not, it is possible that our methods may detect incompatibility when it is not present. For example, the variance inflation parameter may activate to accommodate a summary statistic distribution with a heavy tail, even when the model is correct. We are currently working on adapting our approach to other likelihood-free methods that relax the Gaussian assumption, such as ABC and the semi-parametric extension of BSL An et al. (2020), which uses flexible models for the marginal summary statistic distributions.

The choice of prior for γ was chosen for simplicity and ease of posterior sampling via the slice sampler. We note, however, that other prior choices are possible. For example, it is worth investigating sparsity inducing priors such as spike and slab priors. However, the mixed discrete-continuous nature of these priors may complicate posterior sampling in our context. Another possible choice is the class of global-local shrinkage priors such as the horseshoe. There is also a question of hyperparameter choice, although our default choice is sensible and produced good empirical results in our examples. Attempting to tune the hyperparameter by cross validation or running prior sensitivity studies would be highly computationally intensive in our likelihood-free context. We leave a more thorough investigation of prior choice to future research.

Lastly, we note that, given the similarities between BSL and ABC, a natural question posed during this research was whether or not the mean and variance adjustment approaches discussed in this current paper were applicable in the context of ABC. In concurrent work from the authors, preliminary investigations into a similar type of mean and variance adjusted ABC have revealed that such an approach can mitigate the poor performance of ABC under model misspecification (Frazier et al., 2020).

Acknowledgments

The authors are grateful to Ziwen An, who provided some useful code for this paper. Frazier was supported by the Discovery Early Career Researcher Award funding scheme.

References

- An, Z., Nott, D. J., and Drovandi, C. (2020). Robust Bayesian synthetic likelihood via a semi-parametric approach. *Statistics and Computing*, 30:543557.
- Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725.
- Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2012). Bayesian shrinkage. *arXiv preprint arXiv:1212.6088*.
- Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130.
- Bortot, P., Coles, S. G., and Sisson, S. A. (2007). Inference for stereological extremes. *Journal of the American Statistical Association*, 102(477):84–92.
- Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.
- Browning, A. P., McCue, S. W., Binny, R. N., Plank, M. J., Shah, E. T., and Simpson, M. J. (2018). Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. *Journal of Theoretical Biology*, 437:251–260.
- Cranmer, K., Brehmer, J., and Louppe, G. (2019). The frontier of simulation-based inference. *arXiv preprint arXiv:1911.01429*.
- Fearnhead, P. (2018). Asymptotics of abc. In *Handbook of Approximate Bayesian Computation*, pages 269–288. Chapman and Hall/CRC.
- Frazier, D. T., Martin, G. M., Robert, C. P., and Rousseau, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika*, 105(3):593–607.
- Frazier, D. T., Nott, D. J., Drovandi, C., and Kohn, R. (2019). Bayesian inference using synthetic likelihood: asymptotics and adjustments. *arXiv preprint arXiv:1902.04827*.
- Frazier, D. T., Robert, C. P., and Rousseau, J. (2020). Model misspecification in approximate bayesian computation: consequences and diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2):421–444.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*, 4:641–649.
- Gutmann, M. U. and Corander, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *The Journal of Machine Learning Research*, 17(1):4256–4302.

- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons.
- Kleijn, B. and Van der Vaart, A. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6:354–381.
- Marchand, P., Boenke, M., and Green, D. M. (2017). A stochastic movement model reproduces patterns of site fidelity and long-distance dispersal in a population of Fowler’s toads (*Anaxyrus fowleri*). *Ecological Modelling*, 360:63 – 69.
- Marin, J.-M., Pillai, N. S., Robert, C. P., and Rousseau, J. (2014). Relevant statistics for Bayesian model choice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(5):833–859.
- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.
- Papamakarios, G., Sterratt, D. C., and Murray, I. (2018). Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. *arXiv preprint arXiv:1805.07226*.
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.
- Price, L. F., Drovandi, C. C., Lee, A., and Nott, D. J. (2018). Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 27(1):1–11.
- Ratmann, O., Andrieu, C., Wiuf, C., and Richardson, S. (2009). Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proceedings of the National Academy of Sciences*, 106(26):10576–10581.
- Rieder, H. (2012). *Robust asymptotic statistics*, volume 1. Springer Science & Business Media.
- Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Pettet, G. J. (2015a). Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. *PLOS Computational Biology*, 11(12):e1004635.
- Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Simpson, M. J. (2015b). Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. *Mathematical Biosciences*, 263:133–142.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102.

Appendix

This appendix contains additional details for the contaminated normal example and an empirical example that analyzes the behavior of a popular collective cell spreading model. In addition, this material includes the proof of Proposition 1 in the main text.

A Additional Details: Normal Example

In this section, we present additional details for Example 1 in the paper. First, we analyze the R-BSL adjustment components, Γ , from Example 1, and then we present repeated sampling results for BSL and R-BSL across three levels of model misspecification. We refer the reader to Sections 2.3 and 4.1 of the main paper for details of the Monte Carlo specification.

A.1 Adjustment Components

Figure 12 displays the resulting posterior densities for Γ across the two R-BSL procedures, and across all levels of misspecification. Panels A and B give the results for R-BSL-M, and correspond to the components γ_1 and γ_2 , respectively, while panels C and D give the same results for R-BSL-V. For comparison purposes, the black line in each panel represents the prior densities, and the color-coding in each figure represents the level of misspecification (where $\sigma_\epsilon^2 = 1$ encodes correct model specification).

Focusing on Panel A, we see that the posterior densities for the γ_1 component, in R-BSL-M, which captures our ability to match the first observed statistic (the mean), are indistinguishable from the prior across all the replicated data sets, which implies that we can match the mean of this model regardless of model misspecification. In contrast, in Panel B we see that the second component, which captures our ability to match the second observed statistic (the variance), looks nothing like the prior, except perhaps at low levels of misspecification.

Panels C and D describe precisely the same story as in Panels A and B but correspond to γ_1 and γ_2 in R-BSL-V. Under correct specification ($\sigma_\epsilon^2 = 1$) the posteriors are indistinguishable from the priors, we are easily able to detect departures from compatibility for the second summary statistic, and the posteriors for the first adjustment term remain indistinguishable from the prior across all data sets. These results demonstrate that both R-BSL approaches are capable of reliably detecting which features of the model we are not able to replicate.

A.2 BSL and R-BSL Comparison

We now analyze the repeated sampling behavior of R-BSL and compare it to BSL. We again consider data generated from the contaminated normal model, and we choose σ_ϵ^2 in equation (3) in the main text so that the observed data for \mathbf{y} has *sample variance*

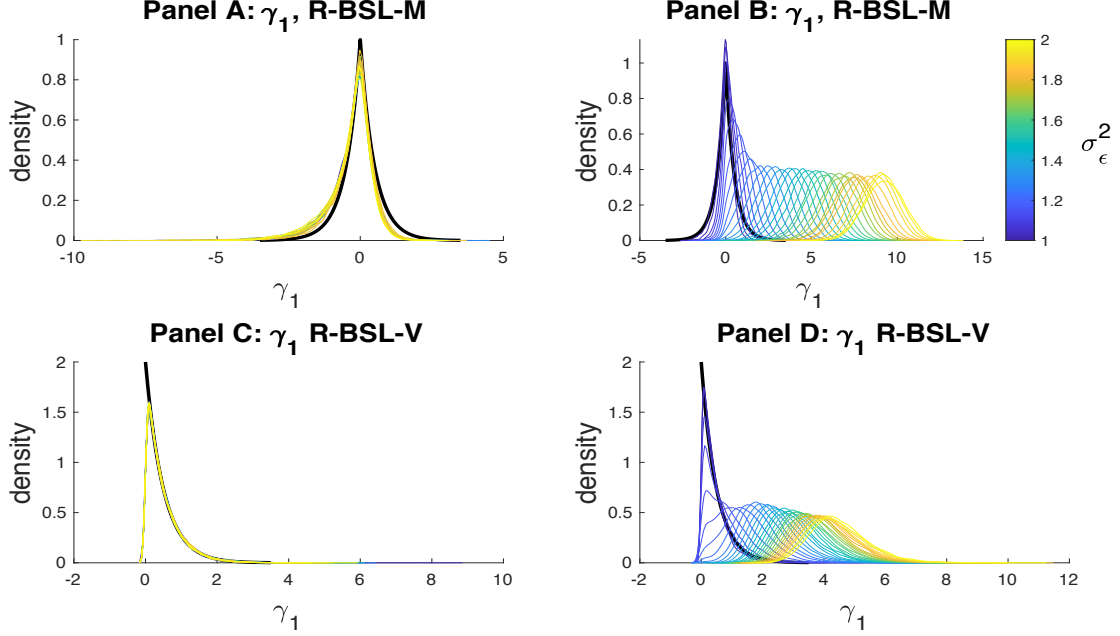


Figure 12: Marginal posteriors for adjustment components. Panels A and B correspond to the mean adjustment R-BSL approach (R-BSL-M), while Panels C and D correspond to the variance adjustment R-BSL approach (R-BSL-V). Panels A and C correspond to the components for the first summary, and panels B and D for the second summary. In each sub-figure, the black line corresponds to the prior density for each component.

equal to $\{1.00, 1.50, 2.00\}$. For each value of the sample variance, we generate one hundred replicated data sets. We then apply R-BSL and BSL, and calculate the bias (BIAS), and root mean squared error (RMSE), associated with the posterior means, as well as the average credible set length (LEN), obtained from a 95% confidence set, and the Monte Carlo coverage (COV), as determined by a 95% confidence set. Table 2 displays the results and demonstrates that R-BSL performs well across both correctly and incorrectly specified models. In the case of correct specification, R-BSL-V and BSL perform very similarly. However, as model misspecification increases, R-BSL-V yields more accurate point estimators than BSL, and better uncertainty quantification. These results demonstrate that when the model is misspecified, R-BSL will yield more reliable statistical inferences than those obtained by BSL.

Table 2: Summary measures for posterior mean accuracy, calculated as averages across the Monte Carlo replications. RMSE- root mean squared error of the posterior mean, BIAS- bias of the posterior mean, LEN- credible set length, COV- Monte Carlo coverage. BSL refers to BSL, R-V refers to R-BSL-V, and R-M refers to R-BSL-M. Average acceptance rates across the three different levels of misspecification are as follows: R-BSL-V: 71.86%, 53.47%, 41.78%; R-BSL-M: 68.44%, 24.51%, 5.83%; and BSL: 68.77%, 3.93%, 0.02%

	$\sigma=1.0$			$\sigma=1.5$			$\sigma=2.0$		
	BSL	R-V	R-M	BSL	R-V	R-M	BSL	R-V	R-M
RMSE	0.0022	0.0034	0.0292	0.0175	0.0029	0.0299	0.1436	0.0026	0.0257
BIAS	0.0000	0.0003	0.0296	-0.0006	0.0006	0.0297	0.0165	0.0003	0.0286
LEN	0.3917	0.4807	0.5065	0.4515	0.4821	0.5120	0.1916	0.4826	0.5185
COV	100%	100%	100%	100%	100%	100%	52%	100%	100%

B Collective Cell Spreading

B.1 Background

Collective cell spreading models are often used to gain insight into the biological mechanisms governing, for example, wound healing and skin cancer growth (e.g. Vo et al. (2015b,a)). Browning et al. (2018) develop a simulation-based model where cells are able to move freely in continuous space. They calibrate the model to real *in vitro* data collected from a cell proliferation assay experiment using a rejection-based ABC algorithm. Here, using our new synthetic likelihood methods, we demonstrate that the model is not compatible with the observed summary statistic and provide insight into what aspects of the data that the model is not able to recover.

The model of Browning et al. (2018) is a stochastic individual based model where cells move and interact in a two dimensional space. Here we provide only brief details of the model and refer to Browning et al. (2018) for the full description. Proliferation (cell birth) and motility (movement) for each cell evolves in continuous time according to a Poisson process. The intrinsic rates are given by p and m for proliferation and motility events, respectively. The rates of these processes are also neighbourhood-dependent, with rates decreasing as the amount of crowding around a cell increases. The closeness of cells is governed by a Gaussian kernel that depends on a fixed cell diameter, σ . When a cell proliferates, it places a new cell randomly in its neighbourhood according to an uncorrelated two dimensional Gaussian centered at the cell location with component variances of σ^2 . When a motility events occurs, the cell moves a distance of σ . The direction of the move depends on cell density, biased towards lower cell density. A parameter used to help determine the move direction, γ_b , is part of a Gaussian kernel used to measure the closeness of cells. The parameter of interest is $\theta = (p, m, \gamma_b)^\top$.

In the experiments of Browning et al. (2018), images of the cell population are taken every 12 hours starting at 0 hours with the final image taken at 36 hours. Browning et al. (2018) use the number of cells and the pair correlation computed from each image as the summary statistics, resulting in a six dimensional summary statistic here. The pair

correlation is the ratio of the number of pairs of agents separated by some pre-specified distance to an expected number of cells separated by the same distance if the cells were uniformly distributed in space.

The prior distribution is set as $p \sim \mathcal{U}(0, 10)$, $m \sim \mathcal{U}(0, 0.2)$ and $\gamma_b \sim \mathcal{U}(0, 20)$ with no dependence amongst parameters, as in Browning et al. (2018). We use MCMC to sample the posterior with 50,000 iterations and no burn-in as we initialise the chain at the point estimate $\theta = (1, 0.04, 6)^\top$ reported in Browning et al. (2018). We sample over the space of a logit-type transformation of θ so that any proposal is within the prior bounds. We use a multivariate Gaussian random walk proposal on the transformed space with a covariance matrix obtained via some pilot MCMC runs. We use $m = 50$ model simulations to estimate the synthetic likelihood at each MCMC iteration. We run our methods on both simulated (using the point estimate of Browning et al. (2018)) and real data.

B.2 Results

For the priors for each component of γ , we use a Laplace distribution with a scale of 0.5 for the mean adjustment method and an exponential prior with mean of 0.5 for the variance inflation method.

Firstly we present results for the simulated data (model correctly specified). As shown in Figures 13 and 14, the posterior distributions on θ are similar regardless of whether BSL or R-BSL is applied. The MCMC acceptance rates for BSL and R-BSL-M are both 21% and 20%, respectively. The variance inflation seems to allow for a slightly increased acceptance rate (24%) compared to mean adjustment.

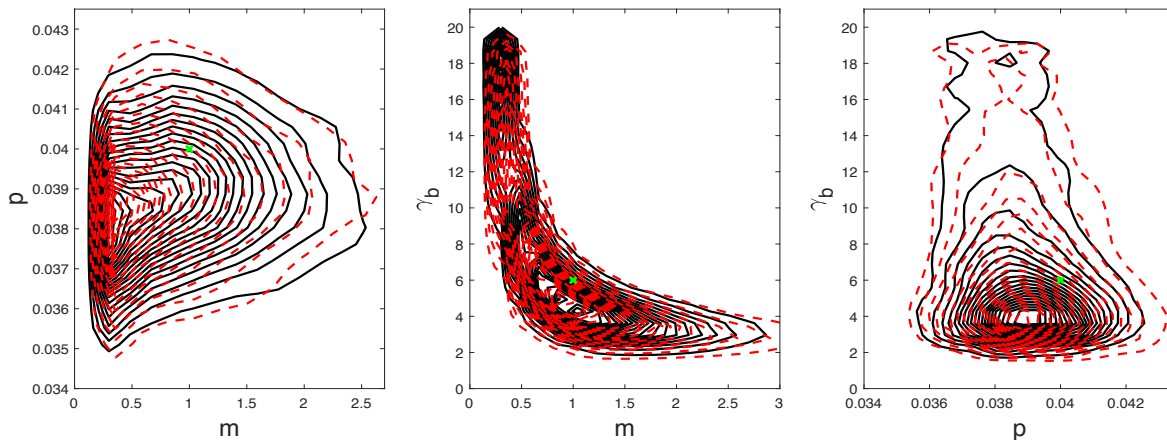


Figure 13: Contour plots of the posterior distributions based on the simulated data for the collective cell spreading example. Results are shown for BSL (solid) and R-BSL-M (dash). The true parameter values are shown as a crosses.

The posterior distribution for each component of γ is shown in Figure 15. It can be seen that most posteriors are similar to the prior. For both R-BSL methods, there is no indication that any of the statistics are incompatible with the model, as expected.

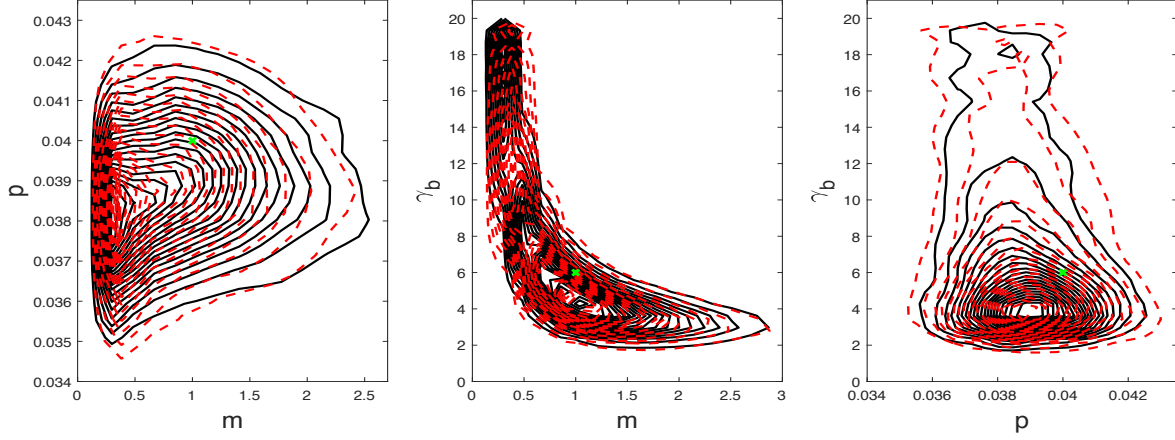


Figure 14: Contour plots of the posterior distributions based on the simulated data for the collective cell spreading example. Results are shown for BSL (solid) and R-BSL-V (dash). The true parameter values are shown as crosses.

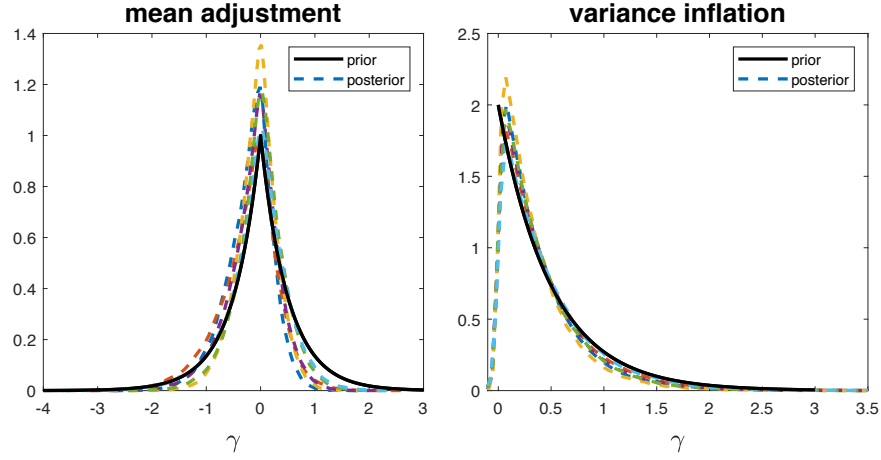


Figure 15: Posterior distributions for each component of γ (dashed lines) based on the simulated data for the collective cell spreading example for R-BSL-M (left) and R-BSL-V (right). The prior distribution of γ is shown as solid lines for both R-BSL-M (left) and R-BSL-V (right).

For the real data, the MCMC acceptance rate using BSL is only 3% as the variance of the synthetic likelihood is high generating long periods of no acceptance. Applying R-BSL-M and R-BSL-V results in an MCMC acceptance rate of roughly 12% and 18%, respectively, permitting statistical inference. Again, the variance inflation seems to produce an improved acceptance rate.

The univariate posterior distributions for γ for R-BSL-M and R-BSL-V are shown in Figures 16 and 17, respectively. Both methods identify that the model is not compatible with the pair correlation statistic at 12 and 36 hours.

In Figure 18, we show R-BSL-M posterior predictive distributions of the summary statistics without (left) and with (right) using estimated mean adjustment parameters

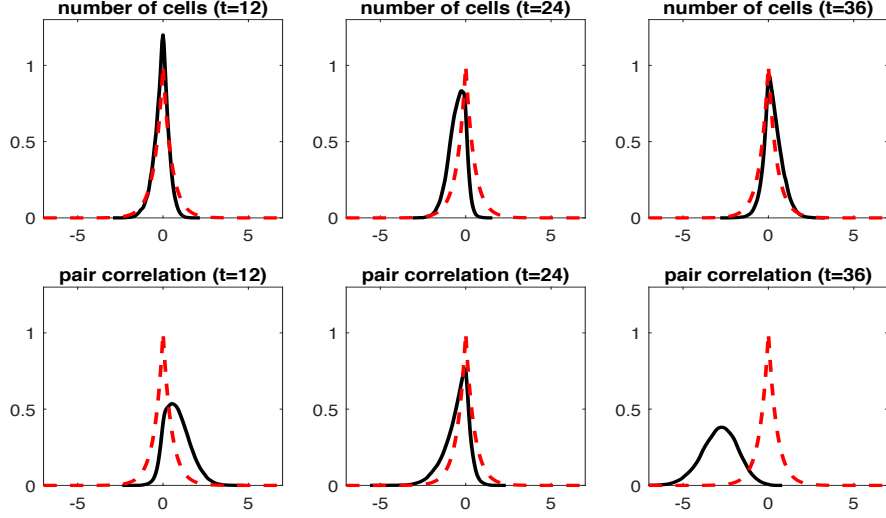


Figure 16: Posterior distributions (solid) for each component of γ when applying R-BSL-M to the real data of the collective cell spreading example. The prior distributions, which are Laplace distributed with scale 0.5, are also shown (dash).

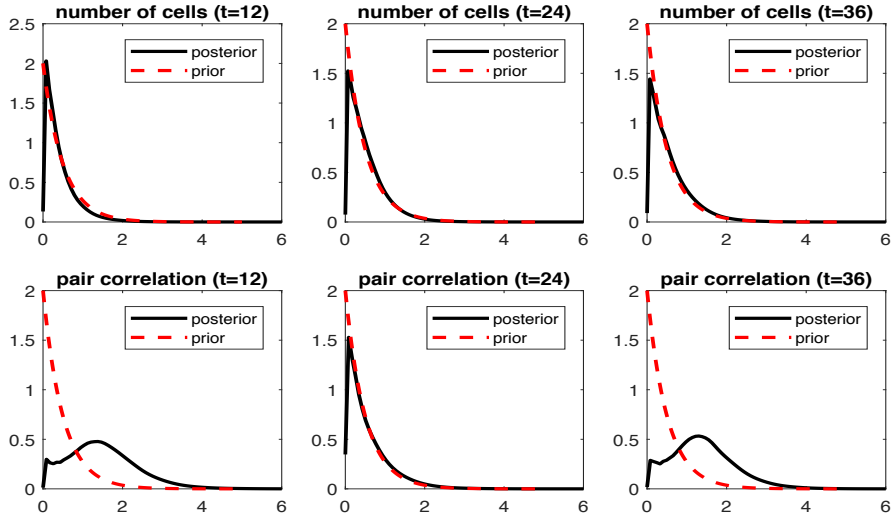


Figure 17: Posterior distributions (solid) for each component of γ when applying R-BSL-V to the real data of the collective cell spreading example. The prior distributions, which are exponential with mean 0.5, are also shown (dash).

with the observed summaries overlaid. The corresponding plots for the variance inflation is shown in Figure 19. From both figures, it is evident from the plots on the left that the model is successful in tracking the number of cells over time. However, the model underestimates the rate of decrease in the pair correlation over time. This is valuable information that might enable mathematical biologists to extend the model so that this data feature can be better captured. Both adjustment methods have allowed us to make this inference. It can

be seen in the second column of Figure 18 that the mean adjustment is able to shift the predictions so that the observed statistic does not lie so far in the tails. From the second column of Figure 19, the variance adjustment expands the predictions so that the observed statistic does not lie so far in the tails.

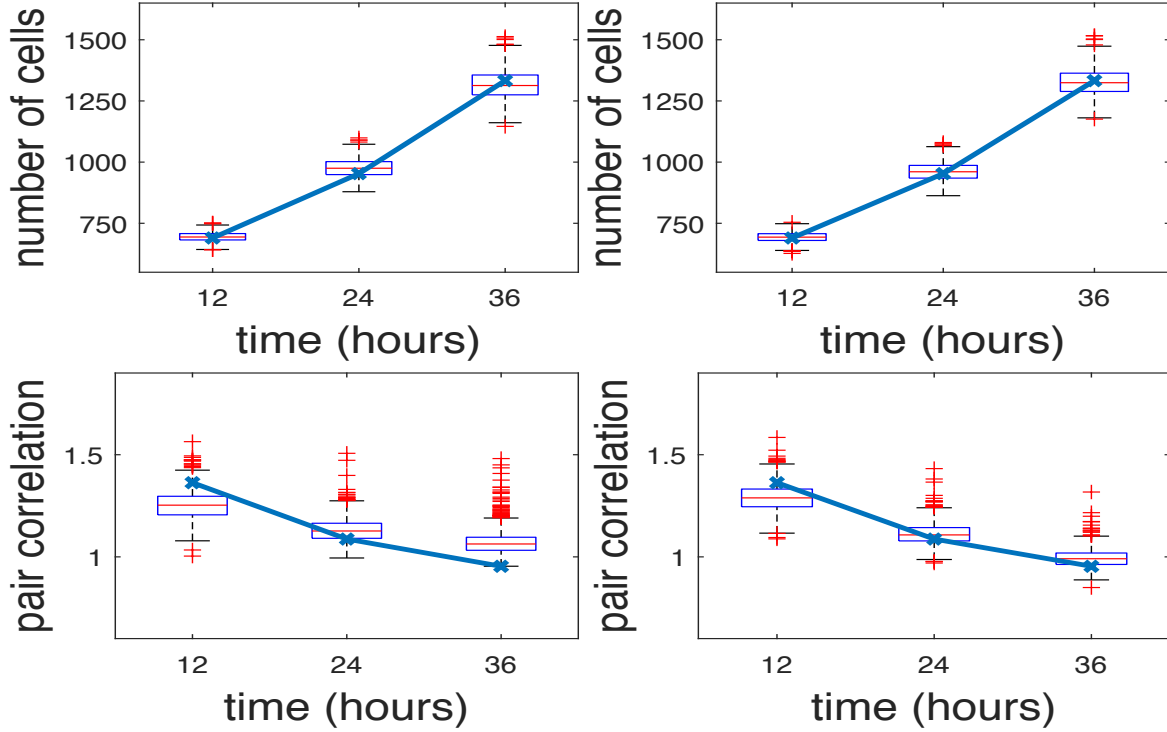


Figure 18: Posterior predictive distributions (summarised as boxplots) of the summary statistics obtained with R-BSL-M. Shown are the posterior predictive distributions of the summaries without (left) and with (right) including the variance inflation in the predictions. The observed summary statistics are overlaid as crosses.

Finally, Figure 20 compares the posterior distributions for the mean adjustment and variance inflation, together the standard synthetic likelihood results. It can be seen that the posterior distributions are broadly similar, except that the standard synthetic likelihood results suffer from substantial Monte Carlo error due to the small acceptance rate.

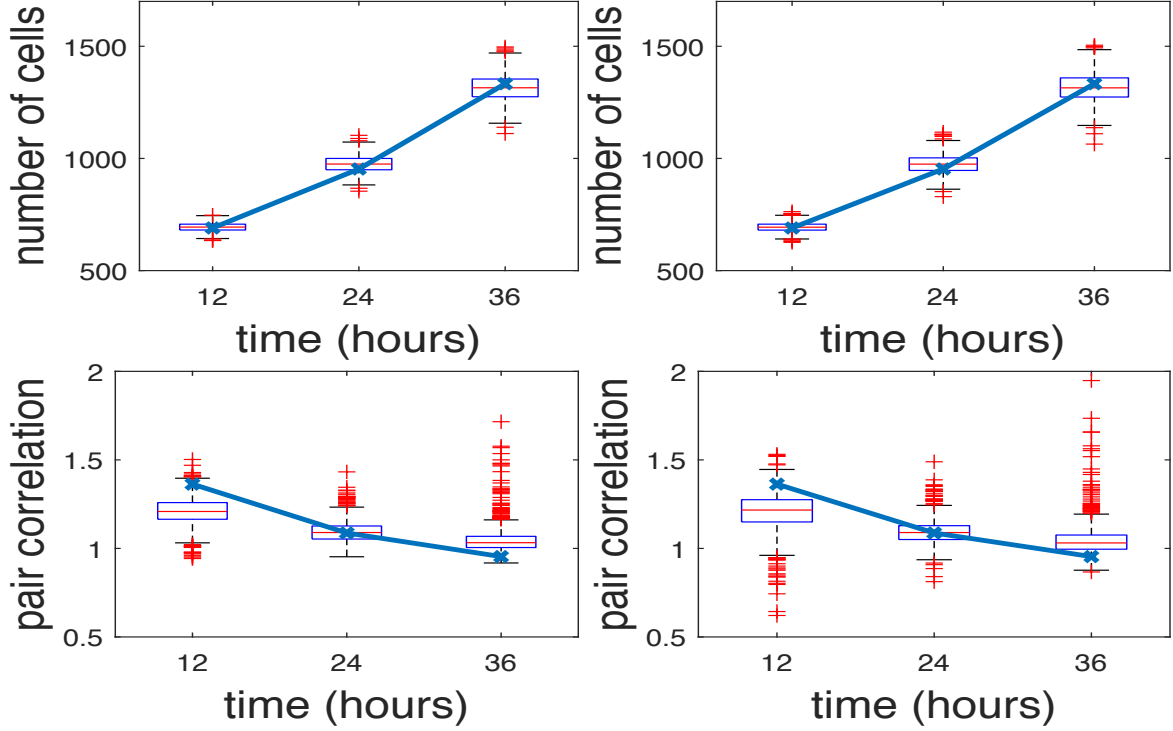


Figure 19: Posterior predictive distributions (summarised as boxplots) of the summary statistics obtained with R-BSL-V. Shown are the posterior predictive distributions of the summaries without (left) and with (right) including the variance inflation in the predictions. The observed summary statistics are overlaid as crosses.

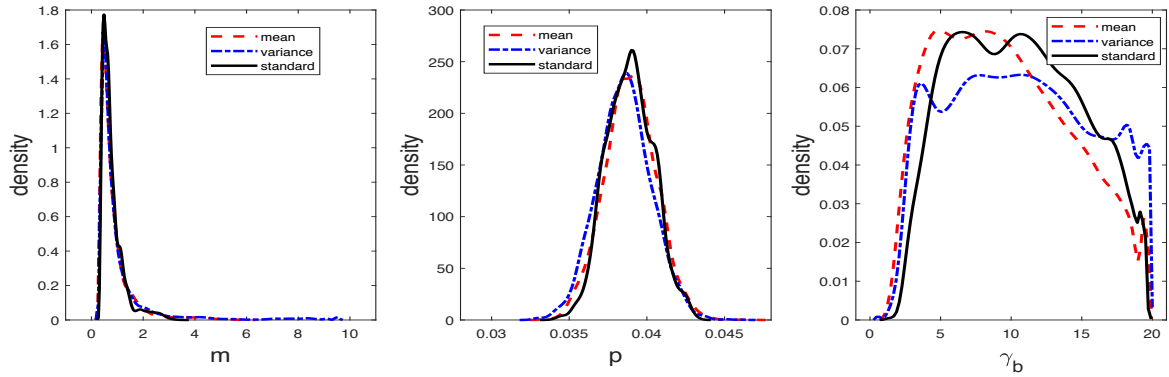


Figure 20: Posterior distributions for each component of θ when applying BSL (solid), R-BSL-M (dash) and R-BSL-V (dot-dash) to the real data of the collective cell spreading example. The priors used are Laplace with scale 0.5 for mean adjustment and exponential with a mean of 0.5 for variance inflation.

C Theoretical Properties of R-BSL

For clarity, we first recall the assumptions and the result in question.

Assumption 7. *There exists a sequence of positive real numbers v_n diverging to ∞ such that, for some distribution Q on \mathbb{R}^{d_η} and some vector $b_0 \in \mathbb{R}^{d_\eta}$,*

$$v_n [\eta(\mathbf{y}) - b_0] \Rightarrow Q, \text{ under } P_n^0.$$

Assumption 8. *(i) The sequence $\{v_n\}_{n \geq 1}$ is such that, for all $\theta \in \Theta$ and some n large enough, there exists constants c_1, c_2 , $c_1 \leq c_2$, such that, for $\|\cdot\|_*$ denoting a matrix norm, $0 < c_1 \leq \|v_n A_n(\theta)\|_* \leq c_2 < \infty$; (ii) For all $\theta \in \Theta$ and all $n \geq 1$, the $d \times d$ -matrix $A_n(\theta)$ is continuous in θ .*

Assumption 9. *There exists a deterministic map $b : \Theta \rightarrow \mathcal{B}$, such that, for all $\theta \in \Theta$, and for constants $\alpha, u_0 > 0$, for all $0 < u < u_0 v_n$,*

$$G_n [\|v_n \{\eta(\mathbf{y}) - b(\theta)\}\| > u \mid \theta] \leq c(\theta) u^{-\alpha},$$

uniformly for $n \geq 1$ and where $\int_{\Theta} c(\theta) \pi(\theta) d\theta = O(1)$.

Assumption 10. *(i) There exists some $\tau > 0$ such that, for all $0 < u < u_0 v_n$, the prior probability satisfies*

$$\Pi [\|b(\theta) - b_0\| \leq u] \asymp u^\tau.$$

(ii) The prior density $\pi(\theta)$ is continuous and satisfies $\pi(\theta_0) > 0$.

Assumption 11. *(i) The map $\theta \mapsto b(\theta)$ is continuous and injective, with $b(\theta_0) = b_0$ for some $\theta_0 \in \Theta$, and satisfies: $\|\theta - \theta_0\| \leq L \|b(\theta) - b_0\|^\kappa$ on some open neighbourhood of θ_0 with $L > 0$ and $\kappa > 0$.*

Assumption 12. *If Assumption 11 is satisfied, for any $\epsilon > 0$, there exists $u, \delta > 0$ and a set V_n such that for all $\theta \in \{\theta : \|b(\theta) - b_0\| \leq u v_n^{-1}\}$*

$$V_n \subset \{\eta \in \mathbb{R}^{d_\eta} : g_n^0(\eta) \lesssim g_n(\eta \mid \theta)\} \text{ where } P_n^0(V_n^c) < \epsilon.$$

Generally speaking, Assumptions 7-12 impart regularity on the summary statistics needed to deduce a posterior concentration result. Assumptions 7, 9 and 12 are similar to those used by Marin et al. (2014) to deduce concentration of posteriors conditioned on summary statistics, while Assumptions 8, and 10 are specific to the analysis of BSL, and have also been used in Frazier et al. (2019). We now discuss each of the assumptions in detail.

Assumption 7 imposes regularity on the observed summary statistics and requires that they satisfy a convergence in distribution result at rate v_n , but does not restrict this distribution to be Gaussian. Assumption 8 is specific to BSL and requires that the BSL variance matrix be well-behaved for all values of $\theta \in \Theta$. This assumption is needed to rule out cases where the BSL variance does not exist or is ill-conditioned (which would be the case, e.g., if the summary statistics are perfectly correlated). Assumption 9 is a condition on the tails of the simulated summary statistics and requires that they have at least a polynomial tail (i.e.,

it requires the existence of at least α moments) uniformly in θ .¹³ Such a condition allows the application of Markov-type inequalities, which are a key ingredient in many posterior concentration results. Assumption 10 is a condition on the tails of the prior used in BSL, and requires that the tails of the prior are not too thick. This condition is satisfied by any prior that admits an exponential moment. Assumption 11 is an identification condition and requires that the simulated summaries are capable of replicating the asymptotic mean of the observed summary statistics. Together with Assumptions 7 and 10, Assumption 11 ensures that the compatibility condition is satisfied. Assumption 11, or a similar variant, has been used in several studies on the asymptotic behavior of approximate Bayesian procedures (see Fearnhead, 2018 for a detailed discussion on this condition). Assumption 12 requires that, for values of θ such that $\|b(\theta) - b_0\|$ is small, up to a universal constant, the assumed model density can be bounded below by the true model density, pointwise, and that the support over which this bound is satisfied has large probability. This condition allows us to link the assumed and true model within the theoretical analysis. This condition is vacuously satisfied if the assumed and true model coincide. However, such a condition would be overly restrictive since compatibility does not require that the assumed model, $g_n(\cdot|\theta)$, coincides with the true model, $g_n^0(\cdot)$, but only that certain moments of the two models agree.

We note that Assumptions 7, 9 and 12 are similar to Assumptions 1, 2 and 4 imposed by Marin et al. (2014) in their analysis of posteriors conditioned on summary statistics, while variants of Assumptions 7-11 have been used by Frazier et al. (2019) in the analysis of the asymptotic properties of BSL. Furthermore, we note that it is trivial to verify Assumptions 7-12 for the contaminated normal example when compatibility is in evidence (i.e., when $\sigma_\epsilon^2 = 1$). In addition, we recall that in the moving average example the model is not compatible, and we have already verified that Assumption 11 is not satisfied. That being said, we note that by following the analysis in Example 1 of Frazier et al. (2018), Assumptions 7-10 can be verified for this example.

The following result gives the theoretical behavior of the R-BSL posterior under the above assumptions. The proof of the result follows.

Proposition 2. *Under Assumption 7-12, for any $\delta > 0$,*

$$\Pi [\|b(\theta) - b_0\| \leq \delta | \eta(\mathbf{y})] = 1 + o_P(1).$$

Moreover, for any $A \subseteq \mathcal{G}$:

$$\Pi [\Gamma \in A | \eta(\mathbf{y})] = \Pi[\Gamma \in A] + o_P(1).$$

We first prove a result that is of independent interest. Namely, we demonstrate that under Assumptions 7-12, the standard BSL posterior concentrates all posterior mass onto the sets of the form $\{b : \|b - b_0\| \lesssim v_n^{-1}\}$. To simplify the computations, we demonstrate this result for the so-called “idealized” BSL posterior, which takes as the mean and variance

¹³A stronger version of this condition has been used in Frazier et al. (2019) to deduce posterior concentration of the BSL posterior.

the infeasible counterparts

$$b(\theta) := \mathbb{E}[\eta(\mathbf{y})|\theta] \text{ and } A_n(\theta) := \left(\mathbb{E} \left[(\eta(\mathbf{y}) - \mathbb{E}[\eta(\mathbf{y})|\theta]) (\eta(\mathbf{y}) - \mathbb{E}[\eta(\mathbf{y})|\theta])^\top | \theta \right] \right)^{1/2}.$$

For $\Sigma(\theta) := A_n(\theta)^\top A_n(\theta)$, the BSL “likelihood” is then given by

$$g_n(\eta|\theta) := (2\pi)^{-\frac{d_\eta}{2}} \det(\Sigma(\theta))^{-1} \exp \left(-\frac{1}{2} \{\eta - b(\theta)\}^\top \Sigma^{-1}(\theta) \{\eta - b(\theta)\} \right).$$

Even though b and Σ can depend on n , we suppress this dependence for notational simplicity.

The following result is a modification of Corollary 1 in Marin et al. (2014), and the proof follows similarly.

Lemma 1. *Under Assumption 7-12, the idealized BSL posterior $\pi\{\cdot|\eta(\mathbf{y})\}$ concentrates at the rate $1/v_n$ onto b_0 , provided that $\alpha > \tau$.*

Proof. Take M_n to be a sequence diverging to ∞ . Define

$$m(\eta) := \int_{\Theta} g_n(\eta|\theta) \pi(\theta) d\theta,$$

and consider the set

$$T_n(M_n) := \{\theta \in \Theta : \|b(\theta) - b_0\| > M_n v_n^{-1}\}.$$

Now, consider the BSL posterior over the set $T_n(M_n)$:

$$\frac{\int_{T_n(M_n)} g_n[\eta|\theta] \pi(\theta) d\theta}{m(\eta)} \equiv \frac{\int_{T_n(M_n)} \frac{g_n[\eta|\theta]}{g_n^0(\eta)} \pi(\theta) d\theta}{m(\eta)/g_n^0(\eta)}.$$

Define

$$N_n := \int_{T_n(M_n)} \frac{g_n[\eta|\theta]}{g_n^0(\eta)} \pi(\theta) d\theta$$

and

$$D_n := m(\eta) / g_n^0(\eta).$$

By Lemma 2, we have that $D_n \gtrsim v_n^{-\tau}$. Moreover, by Lemma 3, we have that $N_n \lesssim M_n^{-\alpha} v_n^{-\alpha}$. Therefore,

$$\Pi\{T_n(M_n) \mid \eta\} = \frac{N_n}{D_n} = o_P(v_n^{-\alpha}/v_n^{-\tau}) = o_P(1).$$

From the above, conclude that

$$\Pi[\|b(\theta) - b_0\| > M_n v_n^{-1} | \eta] = o_P(1).$$

Applying Assumption 11, we have the stated result:

$$\Pi [\|\theta - \theta_0\| > L\{M_n v_n^{-1}\}^\kappa | \eta] = o_P(1).$$

■

Lemma 2. *Under Assumptions 7-12,*

$$\lim_{n \rightarrow \infty} P_n^0 (m(\eta) / g_n^0(\eta) \gtrsim v_n^{-\tau}) = 1.$$

Proof. Fix $\delta > 0$. By Assumption 7, there exists an M_δ such that

$$P_n^0 \{v_n \|\eta - b_0\| > M_\delta\} < \delta.$$

For all $\epsilon > 0$, by Assumption 12, there exists $U_\epsilon, \delta_\epsilon$ such that, for $\eta \in V_n$,

$$\int_{\Theta} g_n(\eta | \theta) \pi(\theta) d\theta \geq \delta_\epsilon g_n^0(\eta) \pi[\mathcal{F}_n(U_\epsilon)], \text{ for } \mathcal{F}_n(u) := \{\theta \in \Theta : \|b(\theta) - b_0\| \leq uv_n^{-1}\}.$$

Apply Assumption 10 to obtain

$$\int_{\Theta} g_n(\eta | \theta) \pi(\theta) d\theta \geq \delta_\epsilon g_n^0(\eta) v_n^{-\tau}.$$

From the definition of V_n , it follows that, for n large enough, there exists some $c(\epsilon)$ such that

$$P_n^0 \left\{ \int_{\Theta} g_n(\eta | \theta) \pi(\theta) d\theta \geq c(\epsilon) g_n^0(\eta) v_n^{-\tau} \right\} \geq 1 - \epsilon.$$

■

Lemma 3. *Under Assumptions 7-12, for the set $T_n(M_n) := \{\theta \in \Theta : \|b(\theta) - b_0\| > M_n v_n^{-1}\}$,*

$$\lim_{n \rightarrow \infty} P_n^0 \left(\int_{T_n(M_n)} \frac{g_n[\eta|\theta]}{g_n^0(\eta)} \pi(\theta) d\theta \lesssim M_n^{-\alpha} v_n^{-\alpha} \right) = 1.$$

Proof. Fix $\delta > 0$. By Assumption 7, there exists some M_δ such that, for some n large enough,

$$P_n^0 \{\|\eta - b_0\| > M_\delta / v_n\} < \delta.$$

On this set consider the joint probability

$$P_n^0 \left\{ \int_{T_n(M_\delta)} g_n(\eta|\theta) \pi(\theta) d\theta > g_n^0(\eta) v_n^{-\tau} \right\}.$$

Applying Markov inequality and Fubini's theorem

$$\begin{aligned}
& P_n^0 \left\{ \int_{T_n(M_\delta)} g_n(\eta|\theta) \pi(\theta) d\theta > g_n^0(\eta) v_n^{-\tau} \right\} \\
& \leq P_n^0 (\|\eta - b_0\| > M_\delta v_n^{-1}) + v_n^\tau \int_{T_n(M_\delta)} \int_{\|\mathbf{S}_n - b(\theta)\| \geq M_\delta} \frac{1}{g_n^0(t)} g_n^0(t) g_n(t|\theta) \pi(\theta) dt d\theta \\
& \leq P_n^0 (\|\eta - b_0\| > M_\delta v_n^{-1}) + v_n^\tau \int_{T_n(M_\delta)} \int_{\|\eta - b(\theta)\| \geq M_\delta} g_n(t|\theta) \pi(\theta) dt d\theta \\
& \leq P_n^0 (\|\eta - b_0\| > M_\delta v_n^{-1}) + v_n^\tau \int_{T_n(M_\delta)} G_n [\|\eta - b(\theta)\| \geq M_\delta] \pi(\theta) d\theta.
\end{aligned}$$

It then follows that, by Assumption 9,

$$\begin{aligned}
& P_n^0 (\|\eta - b_0\| > M_\delta v_n^{-1}) + v_n^\tau \int_{T_n(M_\delta)} G_n [\|\eta - b(\theta)\| \geq M_\delta] \pi(\theta) d\theta \\
& \leq v_n^\tau (M_\delta v_n)^{-\alpha} \int_{\Theta} c(\theta) \pi(\theta) d\theta \lesssim o(v_n^{\tau-\alpha}),
\end{aligned}$$

where, by Assumption 9, $\int_{\Theta} c(\theta) \pi(\theta) d\theta = O(1)$. Conclude that, for any $\delta > 0$,

$$P_n^0 \left\{ \int_{T_n(M_\delta)} g_n(\eta|\theta) \pi(\theta) d\theta > g_n^0(\eta) v_n^{-\tau} \right\} \leq \delta + o(v_n^{\tau-\alpha}) \leq 2\delta$$

for n large enough. ■

Proof of Proposition 2. First, we prove the stated result for the posterior of Γ .

Part (1): $\Pi[\Gamma \in A|\eta(\mathbf{y})] = \Pi[\Gamma \in A] + o_P(1)$. We prove the result for R-BSL-M and R-BSL-V separately.

R-BSL-M: From the posterior concentration of the BSL posterior we have that, for some $M_n \rightarrow \infty$, with $M_n/v_n \rightarrow 0$, for $A \subseteq \mathcal{G}$,

$$\begin{aligned}
& \Pi[A|\eta(\mathbf{y})] \\
& = \frac{\int_A \int_{\|b-b_0\| \leq M_n/v_n} (2\pi)^{-\frac{d\eta}{2}} \det[\Sigma(b)]^{-1/2} e^{(-\frac{1}{2}[b-\eta-x(\Gamma)]^\top [\Sigma(b)]^{-1} [b-\eta-x(\Gamma)])} \pi(b) \pi(\Gamma) db d\Gamma}{\int \int_{\|b-b_0\| \leq M_n/v_n} (2\pi)^{-\frac{d\eta}{2}} \det[\Sigma(b)]^{-1/2} e^{(-\frac{1}{2}[b-\eta-x]^\top [\Sigma(b)]^{-1} [b-\eta-x])} \pi(b) \pi(\Gamma) db d\Gamma} + o_P(1),
\end{aligned}$$

where $x(\Gamma) := \text{diag}[\Sigma(b)]^{1/2} \Gamma$ and $\eta := \eta(\mathbf{y})$.

Define $Z := v_n(b - \eta)$ and consider the change of variables $b \mapsto v_n(b - \eta) + v_n(\eta - b_0) \equiv Z + v_n(\eta - b_0)$, which yields

$$\Pi[A|\eta(\mathbf{y})] = \frac{N_n}{D_n} + o_P(1),$$

where

$$N_n := \int_A \int_{\|Z\| \leq M_n} \frac{e^{(-\frac{1}{2}[Z/v_n - x(\Gamma)]^\top [\Sigma(Z/v_n + \eta)]^{-1} [Z/v_n - x(\Gamma)])}}{(2\pi)^{d/2} \det [\Sigma(Z/v_n + \eta)]^{1/2}} \pi(Z/v_n + \eta) \pi(\Gamma) dZ d\Gamma,$$

and

$$D_n := \int \int_{\|Z\| \leq M_n} \frac{e^{(-\frac{1}{2}[Z/v_n - x(\Gamma)]^\top [\Sigma(Z/v_n + \eta)]^{-1} [Z/v_n - x(\Gamma)])}}{(2\pi)^{\frac{d_\eta}{2}} \det [\Sigma(Z/v_n + \eta)]^{1/2}} \pi(Z/v_n + \eta) \pi(\Gamma) dZ d\Gamma.$$

We now analyze N_n and D_n separately.

Term D_n : By Assumption 7, $P_n^0 \{\|\eta - b_0\| \leq M_\delta/v_n\} \geq 1 - \delta$ for some $M_\delta, \delta > 0$. On this set, from the definition of M_n ,

$$\begin{aligned} \sup_{\|Z\| \leq M_n} \|\pi(Z/v_n + \eta) - \pi(b_0)\| &= o_P(1) \\ \sup_{\|Z\| \leq M_n} \|\Sigma(Z/v_n + \eta) - \Sigma(b_0)\| &= o_P(1). \end{aligned}$$

The first equation follows from continuity of $\pi(\cdot)$ and the second from continuity of $A_n(\theta) \equiv \Sigma^{1/2}(\theta)$. By the dominated convergence theorem,

$$\frac{D_n}{\pi(b_0) \det [\Sigma(b_0)]^{-1/2}} = \int \int_{\|Z\| \leq M_n} (2\pi)^{-\frac{d_\eta}{2}} e^{-\frac{1}{2}[Z/v_n - x(\Gamma)]^\top [\Sigma(b_0)]^{-1} [Z/v_n - x(\Gamma)]} \pi(\Gamma) dZ d\Gamma + o_P(1).$$

Define $\tilde{Z} := [\Sigma^{1/2}(b_0)v_n]^{-1} Z$ and note that, by Assumption 8, $\tilde{Z} := A(b_0)^{-1} Z + o_P(1)$, for some positive definite matrix $A(b_0)$. We have that

$$\frac{D_n}{\pi(b_0) \det [\Sigma(b_0)]^{-1/2}} = \int (2\pi)^{-\frac{d_\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) d\tilde{Z} d\Gamma + o_P(1).$$

Recall the following: for $x, y \in \mathbb{R}^{d_\eta}$,

$$\int_{\mathbb{R}^{d_\eta}} e^{-\frac{1}{2}x^\top x + x^\top y} dx = (2\pi)^{d_\eta/2} e^{y^\top y/2}.$$

From the above fact and the dominated convergence theorem

$$\int_{\|\tilde{Z}\| \leq M_n} e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} d\tilde{Z} \rightarrow (2\pi)^{d_\eta/2} e^{\frac{1}{2}\Gamma^\top \Gamma}.$$

Apply the above and Fubini's Theorem to deduce

$$\frac{D_n}{\pi(b_0) \det [\Sigma(b_0)]^{-1/2}} \rightarrow_p \int \pi(\Gamma) d\Gamma = 1.$$

Term N_n : Apply the same argument as for D_n , to obtain

$$\frac{N_n}{\pi(b_0)\det[\Sigma(b_0)]^{-1/2}} = \int_A (2\pi)^{-\frac{d_\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) dZ d\Gamma \rightarrow_p \int_A \pi(\Gamma) d\Gamma.$$

Conclude that

$$\Pi[A|\eta(\mathbf{y})] = \frac{N_n}{D_n} + o_P(1) = \Pi[A] + o_P(1).$$

R-BSL-V: Define

$$V(b, \Gamma) := \Sigma(b) + \text{diag}[\Sigma^{1/2}(b)] \Gamma \Gamma^\top \text{diag}[\Sigma^{1/2}(b)]^\top$$

From posterior concentration, for some $M_n \rightarrow \infty$, with $M_n/v_n \rightarrow 0$,

$$\begin{aligned} & \Pi[A|\eta] \\ &= \frac{\int_A \int_{\|b-b_0\| \leq M_n/v_n} (2\pi)^{-\frac{d_\eta}{2}} \det[V(b, \Gamma)]^{-1} e^{(-\frac{1}{2}[b-\eta]^\top V(b, \Gamma)^{-1}[b-\eta])} \pi(b) \pi(\Gamma) db d\Gamma}{\int \int_{\|b-b_0\| \leq M_n/v_n} (2\pi)^{-\frac{d_\eta}{2}} \det[V(b, \Gamma)]^{-1} e^{(-\frac{1}{2}[b-\eta]^\top V(b, \Gamma)^{-1}[b-\eta])} \pi(b) \pi(\Gamma) db d\Gamma} + o_P(1), \end{aligned}$$

where $\eta := \eta(\mathbf{y})$.

Similar to the case of R-BSL-M, define $Z := v_n(b - \eta)$ and consider the change of variables $b \mapsto v_n(b - \eta) + v_n(\eta - b_0) \equiv Z + v_n(\eta - b_0)$, which yields, for some $M_n \rightarrow \infty$,

$$\Pi[A|\eta(\mathbf{y})] = \frac{N_n}{D_n} + o_P(1)$$

where

$$N_n := \int_A \int_{\|Z\| \leq M_n} \frac{e^{(-\frac{1}{2}[Z/v_n]^\top V^{-1}(Z/v_n + \eta, \Gamma)[Z/v_n])}}{(2\pi)^{\frac{d_\eta}{2}} \det[V(Z/v_n + \eta, \Gamma)]^{1/2}} \pi(Z/v_n + \eta) \pi(\Gamma) dZ d\Gamma$$

and

$$D_n := \int \int_{\|Z\| \leq M_n} \frac{e^{(-\frac{1}{2}[Z/v_n]^\top V^{-1}(Z/v_n + \eta, \Gamma)[Z/v_n])}}{(2\pi)^{\frac{d_\eta}{2}} \det[V(Z/v_n + \eta, \Gamma)]^{1/2}} \pi(Z/v_n + \eta) \pi(\Gamma) dZ d\Gamma.$$

Again, we analyze N_n and D_n separately.

Term D_n : Similar to the previous result, by Assumption 7, $P_n^0\{\|\eta - b_0\| \leq M_\delta/v_n\} \geq 1 - \delta$

for some $M_\delta, \delta > 0$. Similarly to the case of R-BSL-M,

$$\begin{aligned} \sup_{\|Z\| \leq M_n} \|\pi(Z/v_n + \eta) - \pi(b_0)\| &= o_P(1) \\ \sup_{\|Z\| \leq M_n} \|\Sigma(Z/v_n + \eta) - \Sigma(b_0)\| &= o_P(1). \end{aligned}$$

By Assumption 8, and the dominated convergence theorem,

$$\frac{D_n}{\pi(b_0)v_n^{-d_\eta}} = \int \int_{\|Z\| \leq M_n} (2\pi)^{-\frac{d_\eta}{2}} \det[v_n^2 V(b_0, \Gamma)]^{-1/2} e^{(-\frac{1}{2}[Z/v_n]^\top [V(b_0, \Gamma)]^{-1} [Z/v_n])} \pi(\Gamma) dZ d\Gamma + o_P(1).$$

Define

$$\tilde{Z} := \left[v_n^2 \Sigma(b_0) + \text{diag}[v_n \Sigma^{1/2}(b_0)] \Gamma \Gamma^\top \text{diag}[v_n \Sigma^{1/2}(b_0)]^\top \right]^{-1/2} Z.$$

By Assumption 8, $\tilde{Z} = A(b_0 + \Gamma)^{-1/2} Z + o_P(1)$ and we then obtain

$$\frac{D_n}{\pi(b_0)v_n^{-d_\eta}} = \int (2\pi)^{-\frac{d_\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} \det[A(b_0 + \Gamma)]^{-1/2} e^{(-\frac{1}{2}\tilde{Z}^\top \tilde{Z})} \pi(\Gamma) dZ d\Gamma + o_P(1).$$

From the dominated convergence theorem,

$$(2\pi)^{-d_\eta/2} \int_{\|\tilde{Z}\| \leq M_n} \det[A(b_0 + \Gamma)]^{-1/2} e^{(-\frac{1}{2}\tilde{Z}^\top \tilde{Z})} dZ \rightarrow_p 1.$$

Applying the above and Fubini's Theorem delivers

$$\frac{D_n}{\pi(b_0)v_n^{-d_\eta}} \rightarrow_p \int \pi(\Gamma) d\Gamma = 1.$$

Term N_n : Apply the same argument as for D_n , to deduce that

$$\frac{N_n}{\pi(b_0)v_n^{-d_\eta}} = \int_A (2\pi)^{-\frac{d_\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} \det[A(b_0 + \Gamma)]^{-1/2} e^{(-\frac{1}{2}\tilde{Z}^\top \tilde{Z})} \pi(\Gamma) dZ d\Gamma + o_P(1) \rightarrow_p \int_A \pi(\Gamma) d\Gamma.$$

Conclude that

$$\Pi[A|\eta(\mathbf{y})] = \frac{N_n}{D_n} + o_P(1) = \Pi[A] + o_P(1).$$

From the arbitrary nature of A , the result holds for any $A \subseteq \mathcal{G}$ by the absolute continuity of $\Pi[\cdot]$.

Part (2): $\Pi[\|b(\theta) - b_0\| \leq \delta | \eta(\mathbf{y})] = 1 + o_P(1)$. The result follows along similar lines to those given for **Part (1)**, the posterior for the Γ components, and, hence, we only sketch the result and only for the R-BSL-M case. The equivalent result for R-BSL-V is very similar and hence omitted.

From the posterior concentration of the BSL posterior in Lemma 1 we have that, for

any $\delta > 0$, and some $M_n \rightarrow \infty$, as $n \rightarrow \infty$, such that $\lim_n M_n/v_n = \delta$,

$$\begin{aligned} & \Pi [\|b(\theta) - b_0\| \leq \delta | \eta(\mathbf{y})] \\ &= \frac{\int_{\|b(\theta) - b_0\| \leq \delta} \int (2\pi)^{-\frac{d\eta}{2}} \det [\Sigma(b)]^{-1/2} e^{(-\frac{1}{2}[b-\eta-x(\Gamma)]^\top [\Sigma(b)]^{-1}[b-\eta-x(\Gamma)])} \pi(b) \pi(\Gamma) db d\Gamma}{\int \int (2\pi)^{-\frac{d\eta}{2}} \det [\Sigma(b)]^{-1/2} e^{(-\frac{1}{2}[b-\eta-x]^\top [\Sigma(b)]^{-1}[b-\eta-x])} \pi(b) \pi(\Gamma) db d\Gamma}, \\ &= \frac{\int_{v_n \|b(\theta) - b_0\| \leq M_n} \int (2\pi)^{-\frac{d\eta}{2}} \det [\Sigma(b)]^{-1/2} e^{(-\frac{1}{2}[b-\eta-x(\Gamma)]^\top [\Sigma(b)]^{-1}[b-\eta-x(\Gamma)])} \pi(b) \pi(\Gamma) db d\Gamma}{\int \int_{v_n \|b-b_0\| \leq M_n} (2\pi)^{-\frac{d\eta}{2}} \det [\Sigma(b)]^{-1/2} e^{(-\frac{1}{2}[b-\eta-x]^\top [\Sigma(b)]^{-1}[b-\eta-x])} \pi(b) \pi(\Gamma) db d\Gamma} + o_P(1), \end{aligned}$$

where $x(\Gamma) := \text{diag}[\Sigma(b)]^{1/2} \Gamma$ and $\eta := \eta(\mathbf{y})$.

Define $Z := v_n (b - \eta)$ and consider the change of variables $b \mapsto v_n (b - \eta) + v_n (\eta - b_0) \equiv Z + v_n (\eta - b_0)$, which yields

$$\Pi [\|b(\theta) - b_0\| \leq \delta | \eta(\mathbf{y})] = \frac{N_n}{D_n} + o_P(1),$$

where

$$N_n := \int_{\|Z\| \leq M_n} \int \frac{e^{(-\frac{1}{2}[Z/v_n - x(\Gamma)]^\top [\Sigma(Z/v_n + \eta)]^{-1}[Z/v_n - x(\Gamma)])}}{(2\pi)^{d/2} \det [\Sigma(Z/v_n + \eta)]^{1/2}} \pi(Z/v_n + \eta) \pi(\Gamma) d\Gamma dZ,$$

and

$$D_n := \int_{\|Z\| \leq M_n} \int \frac{e^{(-\frac{1}{2}[Z/v_n - x(\Gamma)]^\top [\Sigma(Z/v_n + \eta)]^{-1}[Z/v_n - x(\Gamma)])}}{(2\pi)^{\frac{d\eta}{2}} \det [\Sigma(Z/v_n + \eta)]^{1/2}} \pi(Z/v_n + \eta) \pi(\Gamma) d\Gamma dZ.$$

We now analyze N_n and D_n separately, starting with D_n .

Term D_n . The argument is identical to that used to prove the same part in **Part (1)** of the result. Therefore, defining $\tilde{Z} := [\Sigma^{1/2}(b_0)v_n]^{-1} Z$, by Assumption 8 $\tilde{Z} := A(b_0)^{-1} Z + o_P(1)$, for some positive definite matrix $A(b_0)$, and it follows that

$$\begin{aligned} \frac{D_n}{\pi(b_0) \det [\Sigma(b_0)]^{-1/2}} &= \int (2\pi)^{-\frac{d\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) d\tilde{Z} d\Gamma + o_P(1) \\ &\rightarrow_p \int \pi(\Gamma) d\Gamma = 1. \end{aligned}$$

Similar to the proof of **Part (1)**, the second line follows from the dominated convergence theorem and Fubini's theorem.

Term N_n . Again, the argument follows similarly to that used for the posterior of the Γ

components. Similar to the case for D_n , we can obtain

$$\begin{aligned}\frac{N_n}{\pi(b_0)\det[\Sigma(b_0)]^{-1/2}} &= (2\pi)^{-\frac{d_\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} \int e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) d\tilde{Z} d\Gamma \\ &= (2\pi)^{-\frac{d_\eta}{2}} \int_{\|\tilde{Z}\| \leq M_n} \int e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) d\tilde{Z} d\Gamma\end{aligned}$$

From the dominated convergence theorem,

$$\frac{N_n}{\pi(b_0)\det[\Sigma(b_0)]^{-1/2}} \xrightarrow{P} (2\pi)^{-\frac{d_\eta}{2}} \int \int e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) d\tilde{Z} d\Gamma,$$

for \tilde{Z} a standard Gaussian vector. Again, recalling that, for $x, y \in \mathbb{R}^{d_\eta}$,

$$\int e^{-\frac{1}{2}x^\top x + x^\top y} dx = (2\pi)^{d_\eta/2} e^{y^\top y/2},$$

we can apply the above and Fubini's theorem to obtain that

$$\frac{N_n}{\pi(b_0)\det[\Sigma(b_0)]^{-1/2}} \xrightarrow{P} (2\pi)^{-\frac{d_\eta}{2}} \int \int e^{-\frac{1}{2}\tilde{Z}^\top \tilde{Z} + \tilde{Z}^\top \Gamma} e^{-\frac{1}{2}\Gamma^\top \Gamma} \pi(\Gamma) d\tilde{Z} d\Gamma = \int \pi(\Gamma) d\Gamma = 1.$$

Putting the two terms together we have that

$$\Pi[\|b(\theta) - b_0\| \leq \delta | \eta(\mathbf{y})] = \frac{N_n}{D_n} + o_P(1),$$

as stated. ■

References

- An, Z., Nott, D. J., and Drovandi, C. (2020). Robust Bayesian synthetic likelihood via a semi-parametric approach. *Statistics and Computing*, 30:543557.
- Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725.
- Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2012). Bayesian shrinkage. *arXiv preprint arXiv:1212.6088*.
- Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130.

- Bortot, P., Coles, S. G., and Sisson, S. A. (2007). Inference for stereological extremes. *Journal of the American Statistical Association*, 102(477):84–92.
- Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.
- Browning, A. P., McCue, S. W., Binny, R. N., Plank, M. J., Shah, E. T., and Simpson, M. J. (2018). Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. *Journal of Theoretical Biology*, 437:251–260.
- Cranmer, K., Brehmer, J., and Louppe, G. (2019). The frontier of simulation-based inference. *arXiv preprint arXiv:1911.01429*.
- Fearnhead, P. (2018). Asymptotics of abc. In *Handbook of Approximate Bayesian Computation*, pages 269–288. Chapman and Hall/CRC.
- Frazier, D. T., Martin, G. M., Robert, C. P., and Rousseau, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika*, 105(3):593–607.
- Frazier, D. T., Nott, D. J., Drovandi, C., and Kohn, R. (2019). Bayesian inference using synthetic likelihood: asymptotics and adjustments. *arXiv preprint arXiv:1902.04827*.
- Frazier, D. T., Robert, C. P., and Rousseau, J. (2020). Model misspecification in approximate bayesian computation: consequences and diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2):421–444.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*, 4:641–649.
- Gutmann, M. U. and Corander, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *The Journal of Machine Learning Research*, 17(1):4256–4302.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons.
- Kleijn, B. and Van der Vaart, A. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6:354–381.
- Marchand, P., Boenke, M., and Green, D. M. (2017). A stochastic movement model reproduces patterns of site fidelity and long-distance dispersal in a population of Fowler’s toads (*Anaxyrus fowleri*). *Ecological Modelling*, 360:63 – 69.
- Marin, J.-M., Pillai, N. S., Robert, C. P., and Rousseau, J. (2014). Relevant statistics for Bayesian model choice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(5):833–859.

- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.
- Papamakarios, G., Sterratt, D. C., and Murray, I. (2018). Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. *arXiv preprint arXiv:1805.07226*.
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.
- Price, L. F., Drovandi, C. C., Lee, A., and Nott, D. J. (2018). Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 27(1):1–11.
- Ratmann, O., Andrieu, C., Wiuf, C., and Richardson, S. (2009). Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proceedings of the National Academy of Sciences*, 106(26):10576–10581.
- Rieder, H. (2012). *Robust asymptotic statistics*, volume 1. Springer Science & Business Media.
- Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Pettet, G. J. (2015a). Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. *PLOS Computational Biology*, 11(12):e1004635.
- Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Simpson, M. J. (2015b). Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. *Mathematical Biosciences*, 263:133–142.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102.