Journal of Statistical Research *https://doi.org/10.47302/jsr.2020540103*
2020, Vol. 54, No. 1, pp. 43-63 *ISSN 0256 - 422 X*

BAYESIAN ANALYSIS OF FUEL ECONOMY EXPERIMENTS

MOHAMMAD LUTFOR RAHMAN*

Institute of Statistical Research and Training (ISRT)
University of Dhaka, Dhaka 1000, Bangladesh

Email: lutfor@isrt.ac.bd

STEVEN G. GILMOUR

King's College London, Strand, London WC2R 2LS, United Kingdom

Email: steven.gilmour@kcl.ac.uk

PETER J. ZEMROCH AND PAULINE R. ZIMAN

Shell Global Solutions (UK), Shell Centre, York Road, London SE1 7NA, United Kingdom

SUMMARY

Statistical analysts can encounter difficulties in obtaining point and interval estimates for fixed effects when sample sizes are small and there are two or more error strata to consider. Standard methods can lead to certain variance components being estimated as zero which often seems contrary to engineering experience and judgement. Shell Global Solutions (UK) has encountered such challenges and is always looking for ways to make its statistical techniques as robust as possible. In this instance, the challenge was to estimate fuel effects and confidence limits from small-sample fuel economy experiments where both test-to-test and day-to-day variation had to be taken into account. Using likelihood-based methods, the experimenters estimated the day-to-day variance component to be zero which was unrealistic. The reason behind this zero estimate is that the data set is not large enough to estimate it reliably. The experimenters were also unsure about the fixed parameter estimates obtained by likelihood methods in linear mixed models. In this paper, we looked for an alternative to compare the likelihood estimates against and found the Bayesian platform to be appropriate. Bayesian methods assuming some non-informative and weakly informative priors enable us to compare the parameter estimates and the variance components. Profile likelihood and bootstrap based methods verified that the Bayesian point and interval estimates were not unreasonable. Also, simulation studies have assessed the quality of likelihood and Bayesian estimates in this study.

Keywords and phrases: Multi-stratum design, Bayesian analysis, mixed model, fuel economy experiment, simulation study.

AMS Classification: 62

* Corresponding author

1 Introduction

In the context of transport, fuel economy refers to the relationship between the distance traveled by an automobile and the amount of fuel consumed. Fuel economy is dependent on a number of factors such as vehicle and fuel properties, driving patterns and loads. For example, vehicle properties might include engine parameters, weight, aerodynamic drag and rolling resistance while fuel properties might be physical, chemical or performance related, e.g. octane. Fuel effects are generally smaller than vehicle, driving or loading effects, so laboratories have to conduct carefully controlled experiments in order to measure the former. These experiments often involve running vehicles for lengthy periods of time with limited opportunities to change fuel due to the time and effort needed to remove all traces of the previous fuel out of the engine. This means that independent sample sizes can be small and a number of time-related variance components come into play.

This work is motivated by challenges in estimating variance components in small experiments, in particular in fuel economy and round robin studies. In order to demonstrate the effectiveness of its fuels, Shell has conducted many detailed testing programmes comparing the performances of different fuels. The data from one of the fuel economy programmes were analyzed by Shell using likelihood-based methods. The likelihood-based methods estimated certain time-related variance components to be zero for some data sets from that programme but not others. The reason behind this zero estimate is that the data set is not large enough to estimate it reliably. The sampling distribution of the estimator of this variance component is highly positively skewed, leading to the likelihood being maximised at zero. The models used here assumed variance components were non-negative; if negative variance components are allowed, they also frequently arise in practice. Accurate estimates of such components were unobtainable in the current study, possibly owing to the limited amount of relevant data. One such data set will be examined in detail in this paper.

As the experiment was complex, time-consuming and labour-intensive and, as the chosen error model can affect estimates, significance levels and confidence intervals for differences between fuels, the experimenters were interested in maximising the robustness of the statistical techniques employed by also analyzing the data using some other statistical methods for confirmatory purposes. The experimenters believed that a Bayesian approach might resolve the problem associated with the zero estimates of variance components by introducing a certain amount of prior information on the parameters. The Bayesian tools used follow those described in the paper by Gilmour and Goos (2009). Thus a Bayesian method was implemented to overcome the difficulties related to variance component estimation and to provide an alternative against which the outputs obtained from likelihood methods could be compared. In addition, simulation studies were carried out to assess, and to compare the quality of point and interval estimates obtained from likelihood and Bayesian methods.

For the implementation of Bayesian methods a freeware statistical program WinBUGS 1.4 has been used (Lunn et al., 2000). Throughout the simulation studies an associated package R2WinBUGS was used to call WinBUGS from R. Basically R2WinBUGS makes use of the batch mode feature and provides tools to call WinBUGS directly after data manipulation in R. After the WinBUGS process had finished, it was possible to work with the results by importing them back into R. For example, essential posterior summaries were saved in R for further processing, which facilitated the presentation of the results of the simulation studies.

The rest of this article has been organized as follows. Section 2 discusses the underlying design and introduces an example data set. Section 3 describes the likelihood analysis of these fuel economy measurements. Section 4 deals with the Bayesian analysis of the same data set. Section 5 discusses alternative profile likelihood methods and confidence intervals. Simulation studies are discussed in section 6, kernel densities in section 7

and conclusions are drawn in section 8. It should be noted that the studies in section 2 to section 4 were the main concern for Shell and the studies in the remaining sections (section 5 to section 7) have been carried out to further elucidate the properties of the methods.

2 Fuel Economy Experiment

Shell commissioned tests to investigate the effect of adding certain components to a base fuel B to produce a test fuel T. The origin and scale of response have been shifted in the data presented in this paper as they are commercially sensitive. This manipulation of the data did not affect the nature of the statistical analysis under investigation. We have compared the performances of T and B by estimating appropriate contrasts.

In the experiment, fuels were tested in order to assess which gave the better fuel economy in a vehicle. The response variable, measured on a continuous scale, was the distance traveled by a vehicle per gallon of fuel burned.

2.1 Underlying Design

Table 1: Underlying design in the fuel economy experiment

Week	Day	Session	Tests
1	1	AM	BBB
		PM	BBB
	OVERNIGHT STAND		
	2	AM	BBB
		FULL FUEL FLUSH	
		PM	BBB
	OVERNIGHT STAND		
	3	AM	BBB
		PM	BBB
INTERVAL OF 4-5 DAYS		FULL FUEL FLUSH	
2	1	AM	BBB
		PM	BBB
	OVERNIGHT STAND		
	2	AM	BBB
		FULL FUEL FLUSH	
		PM	TTT
	OVERNIGHT STAND		
	3	AM	TTT
		PM	TTT

The tests were carried out over a two week period with three days of actual testing each week, as summarized in Table 1. The programme had two test sessions per day, morning and afternoon.

Once the car had been run on fuel T, it would be very difficult to remove traces of that fuel from the vehicle as it was believed that the additional components might have long-term effects. This limited the number of fuel

changes that could practically be made.

Prior engineering judgement and experience suggested that there might be systematic and/or random morning-to-afternoon, day-to-day and week-to-week effects related to the state of the engine and ambient conditions. These effects had to be catered for in the experimental design. Each session (morning or afternoon), consisted of three back-to-back tests on the same fuel. In week 2, the fuel was changed after the morning session on the second day after a suitable fuel flush. To allow a valid comparison, week 1 was used as a control. The same flushing procedure was used after the morning session on the second day even though the fuel was not actually changed.

In our discussion of the design and subsequent analysis, B-T means a change of treatment from base fuel to test fuel while B-B means a dummy change of treatment, i.e. no change of fuel. Some vehicles had a control week (B-B) followed by a test week (B-T) while others had a test week (B-T) followed by a control week (B-B). An extensive flushing procedure was carried out at the end of week 1. For brevity, we will only analyze data from one vehicle in this paper, tested in the order detailed in Table 1.

2.2 Example Data Set

The fuel economy raw data set is presented in the Appendix (Table A1). The experiment was conducted in two weeks shown in the first column. Days were numbered as 1, 2, 3 for the first week and 4, 5, 6 for the second week. There were two sessions - morning and afternoon - containing three trials each (e.g. BBB or TTT). The response variable Y on a continuous scale represents miles traveled by the vehicle per gallon of fuel.

For simplicity, we assume that the three back-to-back tests in each half-day session can be averaged. We thus have two results per car per day, one from the morning and one from the afternoon, giving 12 data points in all as listed in Table A2. This data set, and subsets thereof, will be analyzed in sections 3 to 5 to evaluate fuel effects by estimating appropriate contrasts. The data will also be used to obtain estimates of between and within day variation, thus enabling confidence intervals and significance tests to be derived for fuel effects.

3 Likelihood Methods in Fuel Economy Experiments

The objective of this section is to analyze the fuel economy experimental data by likelihood-based methods to illustrate the problems associated with the estimation of variance components in small data sets. We have chosen three examples, namely likelihood analysis of the contrasts T-B, B2-B1 and (T-B)-(B2-B1) to illustrate the problem.

3.1 Contrast: T-B

To estimate the contrast T-B, which measures the difference in fuel economy between fuels T and B, we first looked at the week 2 data subset, as listed in Table A3. We fitted linear mixed-effect models by residual maximum likelihood (REML) and maximum likelihood (ML) methods using the statistical software R. The REML and ML methods for linear mixed effects are described in Pinheiro and Bates (2000).

Table 2: Linear mixed-effects model fit for (T-B) by REML and ML methods

Method	Effect	Value	SE	DF	t-value	p-value
REML	α (B)	32.195	0.102	2	316.537	0.000
	β_2	1.332	0.144	2	9.264	0.012
	$\hat{\sigma}_b^2$	1.684×10^{-12}				
ML	α (B)	32.195	0.102	2	316.537	0.000
	β_2	1.332	0.144	2	9.264	0.012
	$\hat{\sigma}_b^2$	1.849×10^{-12}				

Model

The mixed model

$$Y_{jkm} = \alpha + \beta_j + \delta_k + \epsilon_{jkm} \tag{3.1}$$

has been used to estimate the contrast T-B, where $Y_{jkm} \sim N(\mu_{jk}, \tau)$ is the response corresponding to the m-th test (m = 1 or 2) on day k (k = 1, 2, 3) corresponding to fuel j (j = 1 or 2) with mean $\mu_{jk} = E(Y_{jkm}|\delta_k) = \alpha + \beta_j + \delta_k$ and precision $\tau = 1/\sigma^2$. We have cited the normal distribution precision parameter τ rather than σ^2 in the above to keep consistency with the Bayesian terminology used in this paper. The parameter α is the intercept, β_j is the fixed effect due to the jth fuel with the constraint $\beta_1 = 0$, δ_k is the day-to-day error term (i.e. random effects due to the kth day) which follows a normal distribution with mean zero and variance σ_b^2, and ϵ_{jkm} is the within-day error term with mean zero and variance σ^2. We can define within-day correlation by $\rho = \sigma_b^2/(\sigma_b^2 + \sigma^2)$ or equivalently $\rho = \tau/(\eta + \tau)$ which can be also be expressed as $\eta = (1 - \rho)\tau/\rho$, where $\eta = 1/\sigma_b^2$.

Table 2 shows the fits of model (1) to the data presented in Table A3 by the REML and ML methods respectively. The between day variance components obtained by REML and ML methods are 1.684×10^{-12} and 1.849×10^{-12} respectively which, to the all intents, are zero. The reason for this zero estimate of variance component could be small sample size of the experiment. The parameter β_2 provides an estimate of the contrast T-B, indicating that test fuel T gives a fuel economy benefit of 1.332 miles/gallon relative to base fuel B. The p-value is 0.012 in each case suggesting that the performance benefit is statistically significant. The robustness of this inference will be checked by Bayesian methods in section 4.1.

3.2 Contrast: B2-B1

In week 1, the full fuel flushing procedure was conducted halfway through day 2 even though the base fuel B was not actually changed. We were interested in checking whether a change in fuel economy was seen after this flush. Therefore we treated the first nine tests (day 1, day 2 am) as fuel B1 and the next nine tests (day 2 pm, day 3) as fuel B2 (although we knew that both B1 and B2 were actually fuel B). Then we analyzed the week 1 data subset from Table A2 using the same methods as were used for the week 2 data in section 3.1.

Table 3: Linear mixed-effects model fit for (B2-B1) by REML and ML methods

	Effect	Value	SE	DF	t-value	p-value
REML	α (B1)	32.088	0.142	2	225.219	0.000
	β_2	-0.366	0.202	2	-1.815	0.211
	σ_b^2	5.275×10^{-12}				
ML	α (B1)	32.088	0.142	2	225.219	0.000
	β_2	-0.366	0.201	2	-1.815	0.211
	σ_b^2	2.199×10^{-12}				

In the next section 3.3, we will study whether switching from B to T is better than switching from B1 to B2 using the complete week 1 and 2 data set.

Mixed Model

To estimate the contrast B2-B1, we consider the mixed linear model

$$Y_{jkm} = \alpha + \beta_j + \delta_k + \epsilon_{jkm}, \tag{3.2}$$

where Y_{jkm} is the response corresponding to the mth test (m = 1 or 2) on day k (k = 1, 2, 3) corresponding to fuel j (j = 1 or 2), α is the intercept, β_j is the effect due to the jth fuel, δ_k is the random effect due to the kth day and ϵ_{jkm} is the error term corresponding to the mth test (m = 1 or 2) of fuel j (j = 1 or 2) on day k (k = 1, 2, 3). Using the week 1 data subset from Table A2, the results from likelihood based methods are given in Table 3, where β_2 corresponds to the contrast B2-B1. The day-to-day variability is again effectively zero and the p-values in the tables indicate that there is no significant difference between the performances of B1 and B2 in consecutive trials.

3.3 Contrast: (T-B)-(B2-B1)

In this section, we investigate whether test fuel T bestows a fuel economy benefit by comparing its performance against base fuel B in week 2 against the corresponding dummy change in control week 1, as planned in the experimental design (see section 2). By studying the contrast (T-B)-(B2-B1), we are trying to minimize the effects of any systematic morning to afternoon differences and/or day-to-day trends on our estimate of the fuel difference. Data regarding contrast (T-B)-(B2-B1) are given in Table A4.

Mixed Model

We have created three dummy variables in the mixed model (3.3)

$$Y_{jkm} = \alpha + \beta_2 D_{2k} + \beta_3 D_{3k} + \beta_4 D_{4k} + \delta_k + \epsilon_{jkm}, \tag{3.3}$$

Table 4: Linear mixed-effects model fit for (T-B)-(B2-B1) by the REML and ML methods

	Effect	Value	SE	DF	t-value	p-value
REML	α (B1)	32.088	0.124	5	259.225	0.000
	β_2	-0.366	0.175	3	-2.089	0.128
	β_3	0.107	0.175	3	0.610	0.585
	β_4	1.439	0.175	3	8.222	0.004
	σ_b^2	1.510×10^{-12}				
ML	α (B1)	32.088	0.124	5	259.225	0.000
	β_2	-0.366	0.175	3	-2.089	0.128
	β_3	0.107	0.175	3	0.610	0.585
	β_4	1.439	0.175	3	8.222	0.004
	σ_b^2	2.312×10^{-12}				

where Y_{jkm} is the response corresponding to fuel j ($j = 1, 2, 3, 4$) on day k ($k = 1, 2, \ldots, 6$), α is the intercept, β_j is the effect due to the jth fuel (note that $j = 2, 3,$ and 4 correspond to fuels B2, B and T respectively), D_{2k} is the dummy variable corresponding to fuel B2, D_{3k} is the dummy variable corresponding to fuel B, D_{4k} is the dummy variable corresponding to fuel T, δ_k is the random effect due to the kth day and ϵ_{jkm} is the error term corresponding to the mth test (m = 1 or 2) of fuel j ($j = 1, 2, 3, 4$) on day k ($k = 1, 2, \ldots, 6$). Thus we have the means α, $\alpha + \beta_2$, $\alpha + \beta_3$, and $\alpha + \beta_4$ corresponding to fuels B1, B2, B and T respectively.

The REML and ML estimates of the fixed effects in model (3.3) are shown in Table 4 and are similar. We need the estimate of the contrast (T-B)-(B2-B1) to compare with the Bayesian counterparts. The estimates of the contrast (T-B)-(B2-B1) can be obtained from $\{(\alpha + \beta_4) - (\alpha + \beta_3) - (\alpha + \beta_2) + \alpha\}$ which simplifies to $\beta_4 - \beta_3 - \beta_2$. Thus the estimate of contrast (T-B)-(B2-B1) is 1.698 (using either REML or ML estimates from Table 4). Its S.E. is 0.247 with 3 d.f., a t-value of 6.861 and a p-value of 0.006. This again suggests that test fuel T gives significantly better fuel economy than base fuel B.

The estimates for the between days variance $\hat{\sigma}_b^2$ are 1.510×10^{-12} and 2.312×10^{-12} in the REML and ML methods respectively, implying that the between days variance component is effectively zero which again is considered unrealistic.

It is evident that likelihood based methods provide estimates of the day-to-day variance component which are zero or very close to zero. As a consequence, the estimates of the contrasts T-B, B2-B1 and (T-B)-(B2-B1) derived by ML and REML methods in sections 3.1 to 3.3, and the mean values for T, B, B2 and B1 derived therefrom, are equal to the corresponding values obtained from the arithmetic means of the data in Table A4.

In the first instance, Shell used the analysis in section 3.3 to determine whether there were significant fuel differences or not; the Bayesian analysis in section 4.3 was subsequently used as a confirmation of the robustness of the overall approach.

Table 5: Bayesian model fit for contrast T-B

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
B (α)	32.170	0.205	0.002	31.780	32.170	32.580
Fuel Diff(β_2)	1.421	0.240	0.002	0.952	1.420	1.877
T	33.590	0.209	0.002	33.190	33.580	34.040
σ_b^2	0.059	0.181	0.003	0.001	0.021	0.362

4 Bayesian Analysis of Fuel Economy Experiments

To overcome the problems associated with classical methods we have performed Bayesian analyses of the fuel economy data. The various contrasts were estimated using the same models, data sets and subsets used in the likelihood analyses in Section 3.

4.1 Contrast: T-B

We begin by estimating the contrast T-B from the week 2 data in Table A3 in order to compare fuels B and T combining both between day and within day information.

Priors and Results

We used WinBUGS 1.4 to fit model (3.1) to the data by Bayesian methods assuming the following priors

$$\alpha \sim N(32, 0.1), \quad \beta_j \sim N(0, 0.001)$$
$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20)$$

The prior for α was centered at 32 to incorporate the notion of the mean fuel effect and was based on the simple arithmetic mean of the data. Therefore, we assume a weakly informative prior for α by taking $\alpha \sim N(32, 0.1)$. Congdon (2007) suggested that, in the absence of prior information about the direction or magnitude of covariate effects, flat priors may be used by taking univariate normal distributions with mean zero and large variance. The effect of using normal priors with means 0 and large variances is that parameter estimates are smoothed towards zero as large variances are used (Galindo-Garre et al., 2004). Therefore, we tried a non-informative prior for β_j by setting $\beta_2 \sim \text{N}(0, 0.001)$ i.e. β_2 follows a normal distribution with mean 0 and low precision 0.001 or large variance 1000. Also, we assumed a non-informative prior for σ by assuming $\log(\sigma) \sim U(-20, 20)$. The beta(1,1) prior for the intra-class correlation ρ is also non-informative and this is used to compute the precision of the day-to-day error term.

The results of the Bayesian analysis are presented in Table 5 assuming the above set of priors. However, a completely non-informative prior for α, namely $\alpha \sim N(0, 0.001)$, provided similar results to those presented in Table 5 with the weakly informative prior $\alpha \sim N(32, 0.1)$. The prior $\alpha \sim N(0, 0.001)$ is considered non-informative as it has very low precision. The default Markov chain Monte Carlo method used by WinBUGS is Gibbs sampling and this proved adequate for this problem. There were some autocorrelation effects in the results before thinning (where thinning refers to removal of some values from the chain). When data

were thinned by a factor 15 (i.e. instead of using every step in the chain, we only used every 15th step), the autocorrelation disappeared. Posterior means were calculated on the basis of a sample size of 10000. The first 1000 samples were ignored to remove initial fluctuations of the chains.

In Table 5 the first column shows the names of effects. The 'Mean' column shows the magnitude of any effect, the columns headed by 'SD' and 'MC Error' denote standard deviations of effects and Monte Carlo errors respectively; P(2.5), Median, and P(97.5) display the 2.5th, median and 97.5th percentiles of the posterior estimates respectively.

The estimated fuel economy for base fuel B is 32.170 miles/gallon and for the test fuel T it is 33.590 miles/gallon. The difference of effects β_2 between test and base fuel is 1.421, which is slightly higher than the 1.332 found by ML methods in section 3.1. Its SE is also higher at 0.240 vs 0.144 but the difference is still statistically significant as the 95% Bayesian credible interval (0.952, 1.877) does not contain zero.

The variance between days σ_b^2 in the Bayesian analysis is 0.059. However, the lower limit of the 95% credible interval of σ_b^2 is very close to zero. As the mean is larger than the median of the distribution of σ_b^2, it implies that the distribution of σ_b^2 is positively skewed which will also be evident in the portrayal of the kernel density of σ_b^2 presented in section 7.

Comparing estimates of the variance component $\hat{\sigma}_b^2$ in Tables 2 and 5, we see that a poorly estimated variance component in likelihood-based methods becomes estimable in the Bayesian method assuming some priors. The quality of the Bayesian credible interval for $\hat{\sigma}_b^2$ will be assessed by comparison with profile likelihood and bootstrap based confidence intervals derived from likelihood methods (section 5) and by simulation studies (section 6).

4.2 Contrast: B2-B1

Previously we considered the contrast B2-B1 in likelihood methods in section 3.2. Now we will consider a Bayesian approach.

Mixed Model

We reconsider the mixed linear model (3.2) regarding the contrast B2-B1 in the Bayesian context.

Priors and Results

We assume the following priors for the parameters in model (3.2)

$$\alpha \sim N(0, 0.001), \quad \beta_2 \sim N(0, 0.001)$$
$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20).$$

The priors for α, β_2, ρ (intra-class correlation), and $\log(\sigma)$ are assumed to be non-informative. Previously, a weakly informative prior for α was considered, namely $\alpha \sim N(32, 0.1)$. As there are no substantial differences in the results either assuming $\alpha \sim N(32, 0.1)$ or $\alpha \sim N(0, 0.001)$ i.e. a non-informative prior for α, we use only non-informative priors for α in the subsequent analysis. The results concerning the contrast B2-B1 are presented in Table 6.

Table 6 shows that there is negligible difference (β_2) in performance between B1 and B2. The difference at -0.411 is slightly larger than the -0.366 obtained by ML methods in section 3.2 and the SE, once again, is

Table 6: Bayesian model fit for contrast B2-B1

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
B1	32.140	0.311	0.003	31.600	32.120	32.770
B2	31.730	0.289	0.003	31.190	31.730	32.330
Fuel Diff β_2	-0.411	0.361	0.004	-1.118	-0.403	0.257
σ_b^2	0.115	0.454	0.005	0.005	0.047	0.562

larger at 0.361 vs 0.201. But the 95% Bayesian credible interval for the fuel difference contains zero so again the difference is not statistically significant.

4.3 Contrast: (T-B)-(B2-B1)

In this section we use the approach described in section 3.3 to test whether fuel T bestows a fuel economy benefit relative to base fuel B, but now use Bayesian methods to estimate the contrast (T-B)-(B2-B1).

Priors and Results

We assume the following priors for the parameters in model (3.3)

$$\alpha \sim N(0, 0.0001), \quad \beta_j \sim N(0, 0.0001), \quad j = 2, 3, 4;$$
$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20).$$

From the analysis summarized in Table 7 we see that the difference between B2 and B1 is not significant as the 95% credible interval for B2-B1 (-0.802, 0.080) includes zero. However, the credible interval for (T-B)-(B2-B1) does not include zero which implies that there is clear evidence of a benefit by switching from B to T rather than switching from B1 to B2. In other words, there is a clear advantage in changing from base fuel B to test fuel T rather than retaining the same base fuel in consecutive trials.

The Bayesian estimate of the contrast (T-B)-(B2-B1) is 1.776, which is slightly higher than the 1.698 by ML methods in section 3.3, and its SE is larger at 0.320 vs 0.247. But both methods indicate a statistically significant effect.

5 Profile Likelihood and Confidence Intervals

As the fuel economy experiment was a small sample experiment, the estimates, and particularly confidence intervals, in the likelihood method which assumes asymptotic normality of estimates might not be precise due to poor estimates of sampling variance. Therefore, we intend to compare estimates from alternative methods, namely profile likelihood and bootstrap, against the Bayesian method. We have computed confidence intervals based on the Wald procedure, profile likelihood and bootstrap methods for the parameters of the model (3.1), based on the week 2 data in Table A3, and compared these with the Bayesian credible intervals.

Table 7: Bayesian model fit for contrast (T-B)-(B2-B1)

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
α (B1)	32.090	0.187	0.004	31.710	32.090	32.460
β_2	-0.365	0.222	0.004	-0.802	-0.371	0.080
β_3	0.063	0.259	0.002	-0.473	0.069	0.562
β_4	1.469	0.259	0.002	0.952	1.467	1.998
B2	31.720	0.173	0.003	31.380	31.720	32.070
B	32.150	0.179	0.004	31.790	32.160	32.500
T	33.560	0.180	0.005	33.210	33.560	33.940
B2-B1	-0.365	0.222	0.004	-0.802	-0.371	0.080
T-B	1.411	0.232	0.006	0.968	1.409	1.912
(T-B)-(B2-B1)	1.776	0.320	0.008	1.138	1.778	2.424
σ_b^2	0.033	0.052	0.001	0.001	0.017	0.177

Table 8: Likelihood and Bayesian estimates with 95% confidence/credible intervals under different methods

	Likelihood Method				Bayesian Method	
		95% CI				95% CI
Parameter	Estimate	Wald	Profile	Bootstrap	Estimate	
α	32.195	(32.032, 32.358)	(31.915, 32.388)	(32.019, 32.365)	32.170	(31.780, 32.580)
β_2	1.332	(1.102, 1.563)	(1.060, 1.713)	(1.067, 1.598)	1.421	(0.952, 1.877)
σ_b^2	0.000	-	(0.000, 0.099)	(0.000, 0.026)	0.059	(0.001, 0.362)

Table 8 shows that the Bayesian estimate of the fuel difference β_2 is 1.421 which differs from its classical counterpart 1.332. Also, the Bayesian credible interval for β_2 is conservative (wider) in comparison to profile likelihood or bootstrap confidence intervals. The main idea behind the implementation of profile likelihood and bootstrap based methods, along with the Bayesian technique, is to observe the differences in point and interval estimates of the variance component σ_b^2. The point estimate of the day-to-day variance component σ_b^2 is zero in the classical method which is assumed unrealistic, whereas the Bayesian estimate is nonzero (0.059). In the profile and bootstrap methods, the lower limit of the 95% confidence interval for σ_b^2 is zero, whereas the lower limit of the 95% Bayesian credible interval is nonzero, but close to zero. However, Lambert et al. (2005) notice that if the variance parameter is close to the boundary at zero, MCMC results tend to produce upwardly biased estimates of variance parameters when inferences are based on the posterior mean. Therefore considering all the figures in Table 8, either point or interval estimates, we conclude that the day to day variance component σ_b^2 might be nonzero and in between likelihood and Bayesian point estimates in the fuel economy experiment. A simulation study will examine the performance of classical and Bayesian methods in estimating the variance component σ_b^2 in section 6.

To conclude this section we summarize that the point estimates of fixed effects do not differ substantially, but Bayesian confidence intervals are wider than the corresponding likelihood intervals. The estimate of the variance component might not be zero, and could lie between 0 and the Bayesian point estimate 0.059. The comparatively larger width of the Bayesian 95% credible intervals could be minimized with an appropriate choice of priors as non-informative priors are likely to give rise to wider credible intervals.

6 Simulation Studies

In this section we present the results of simulation studies of fuel economy experiments to observe the performances of point estimators and confidence/credible intervals under likelihood and Bayesian methods.

In the actual experiments, the fuels were tested in a number of cars, each of which was tested over six days under carefully controlled conditions. In sections 3.1, 4.1 and 5, we estimated the contrast T-B using data from the second week for one of the vehicles, as shown in Table A3. To see how the results of such an experiment might have turned out had it been repeated, a simulation study has been conducted assuming the same test order. Thus we assumed that n=6 tests would be conducted over three days with two tests per day. The first three tests would be on fuel B and the last three tests on fuel T.

In the first of the simulation studies, we generated data using model (3.1) with the random terms δ_k and ϵ_{jkm} each following normal distributions. The parameters were arbitrarily set as $\alpha = 32$, β_2=1.4, σ_b^2=0.05 and σ^2=0.05. These values are close to those actually found in the Bayesian analysis of the real data (see Table 5). Taking into account both practical considerations and the recommendations of Rahman (2015), the simulation size was fixed at 2000.

The simulation results are summarized in Table 9. The first column lists the parameters, the second column presents the bias of the parameter estimates, the third column is the percent relative bias, the fourth column is the root mean squared error (RMSE), the seventh column represents the root median squared error (RMdSE) and rest of the columns are self explanatory. In the table, the first set of results (first three rows) shows the averages of the likelihood estimates, obtained through REML procedures, of the parameters α, β_2, and σ_b^2 computed from 2000 simulations. The results from those simulations where the REML estimation procedure

in R failed, for any reason, are excluded from these averages.

For the day-to-day variance component σ_b^2, we see that the bias and relative bias of the average (mean) of the simulated estimates is larger than the corresponding bias in the simulated median. This happens because mean estimates are affected by unusual (extreme) estimates that arose during the simulations. But the estimation of random effects was not our primary interest, the main concern being to estimate fixed effects, particularly, the difference in the effects of the test and base fuels (β_2). The biases in the mean and median estimates of α and β_2 were both very small. However the coverage probability for β_2 was 0.929 which is significantly below the acceptable range, because, by definition, a 95% confidence interval should have a coverage probability of at least 0.95. However, even if the true coverage probability equals 95%, the coverage probability obtained from a simulation study might not be exactly equal to 0.95 because of the MC error (Galindo-Garre et al., 2004). This error tends to zero when number of simulations tends to infinity. As we implemented 2000 simulations, the MC error was equal to $(0.95 \times 0.05/2000)^{\frac{1}{2}} = 0.00487$ which means that coverage probabilities between 0.9451 and 0.9549 are in agreement with the nominal level of 95%. Therefore, the coverage probability 0.929, corresponding to β_2 in likelihood method, is clearly beyond the range above.

The last six blocks of three rows in Table 9 show the average estimates from 2000 simulated data sets using the Bayesian analysis described in section 4.1 with the priors $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$ and $\log(\sigma) \sim U(-20, 20)$ and various non-informative and weakly informative priors for either ρ or $1/\sigma_b^2$. Though mean based bias and relative bias for the random effect estimate seem less in the likelihood method than the corresponding Bayesian estimates, MLE is unacceptable as the average width of the 95% confidence interval (CI) for σ_b^2 is infinity. This happens because the covariance matrix is non-positive definite in a number of simulations of small sample experiments (here $n = 6$) which causes the upper CI limit to be infinity.

The problems in fitting mixed models by MLE methods discussed above lead us towards choosing Bayesian methods, assuming non-informative or weakly informative priors. It is relevant to discuss how we have chosen the priors in this study. During the selection of non-informative priors, we followed Lambert et al. (2005) who used 13 different non-informative priors in a simulation study that demonstrates the potential influence of using prior distributions believed to be vague. A few of their non-informative or weakly informative priors resemble ours. For fixed effects, we used non-informative normal priors, as we did in our single real study. However, not all of their priors are suitable for studying the variance components in our models. For example, Pareto, half normal, uniform or normal priors produced highly biased estimates and also had low coverage probabilities. Therefore these were not considered further.

The number of Bayesian iterations was 5000 in the analysis of each simulated data set; the thinning factor was 10, the burn-in period 2000, and the number of chains was 4. For the width of Bayesian credible intervals, we simply average the widths of the 95% credible intervals obtained from each of the simulations. The median width is the median of the respective widths of the 95% CIs both in likelihood and Bayesian methods. Median square error (MdSE) was calculated as the median of $(\hat{\beta} - \beta)^2$ (Galindo-Garre et al., 2004). For confidence/credible intervals, we report coverage probabilities which represent the proportion of times that the true parameter lies within the 95% confidence intervals in the likelihood method whereas in the Bayesian case it is the proportion of times that the true parameter lies within 95% Bayesian credible intervals.

Table 9 shows that the use of the completely non-informative prior $\rho \sim$ Beta$(1, 1)$, which is equivalent to $\rho \sim$ U(0,1), for random effects leads to biased estimates thereof. This prior should not be used to estimate the variance component σ_b^2 as the simulation results show that the relative bias is very high at 223.83%. The bias

Table 9: Simulated performance of maximum likelihood and Bayesian estimates in fuel economy experiments assuming true parameter values as $\alpha = 32, \beta_2 = 1.4$, $\sigma_b^2 = 0.05$, σ^2=0.05; priors as $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$, $\log(\sigma) \sim U(-20, 20)$; and sample size $n = 6$

		Mean/Posterior Mean			Median/Posterior Median					
	Parameter		% Relative			% Relative		Coverage	Average width	Median width
		Bias	Bias	RMSE	Bias	Bias	RMdSE	Probability	of 95% CI	of 95% CI
MLE	α	0.004	0.01	0.217	0.002	0.01	0.144	0.971	1.617	1.514
	β_2	-0.005	-0.36	0.284	-0.008	-0.60	0.187	0.929	1.559	1.520
	σ_b^2	0.028	56.08	0.103	-0.008	-16.49	0.048	0.979	∞	1.896
$\rho \sim$ Beta(1, 1)	α	0.059	0.18	0.215	0.037	0.12	0.208	0.957	1.151	1.110
	β_2	-0.033	-2.38	0.253	-0.024	-1.70	0.255	0.939	1.242	1.198
	σ_b^2	0.112	223.83	0.176	0.006	11.20	0.047	0.991	0.853	0.683
$\rho \sim$ Beta(1.5, 1.5)	α	0.058	0.18	0.209	0.039	0.12	0.203	0.954	1.112	1.071
	β_2	-0.038	-2.72	0.248	-0.028	-1.96	0.248	0.943	1.243	1.224
	σ_b^2	0.088	176.89	0.141	0.007	14.36	0.045	0.988	0.719	0.576
$\rho \sim$ Beta(2.5, 2.5)	α	0.045	0.14	0.205	0.029	0.09	0.201	0.953	1.103	1.060
	β_2	-0.032	-2.28	0.247	-0.022	-1.56	0.246	0.948	1.274	1.226
	σ_b^2	0.074	148.40	0.121	0.009	17.54	0.045	0.987	0.583	0.466
$\rho \sim$ Beta(3, 3)	α	-0.004	-0.01	0.203	-0.005	-0.01	0.204	0.953	1.040	1.000
	β_2	-0.002	-0.13	0.247	-0.001	-0.06	0.247	0.941	1.244	1.200
	σ_b^2	0.062	123.66	0.101	0.009	18.34	0.044	0.980	0.513	0.415
$1/\sigma_b^2 \sim$ Gamma(0.01, 0.001)	α	0.234	0.731	0.303	0.140	0.438	0.242	1	2.807	2.770
	β_2	-0.150	-10.687	0.294	-0.131	-9.391	0.287	1	3.170	3.140
	σ_b^2	0.550	1100.205	0.570	-0.013	-26.582	0.017	1	4.406	3.889
$1/\sigma_b^2 \sim$ Gamma(0.1, 0.1)	α	0.347	1.08	0.392	0.258	0.81	0.318	1.000	2.904	2.772
	β_2	-0.200	-14.32	0.311	-0.171	-12.18	0.298	1.000	3.109	3.075
	σ_b^2	1.920	3840.10	1.959	0.321	641.96	0.324	0.941	8.702	7.181

reduces steadily as the weakly informative priors Beta (1.5, 1.5), Beta (2.5, 2.5) and Beta (3, 3) are introduced but still remains large. This is not unusual. Litière et al. (2008) showed that estimates of the variability of random effects are always biased, though the biases induced in the fixed effect parameters are small so long as the variability underlying the random effects distribution is small. Note that the bias in σ_b^2 is much smaller when bias is measured relative to the median rather than the mean.

Table 9 shows that the estimates of fixed effects are fairly close to the true parameter values in both likelihood and Bayesian methods as the bias and relative bias are small. The bias in fixed effects is generally smaller when measured relative to the median rather than the mean, but not hugely so. When the priors $\rho \sim$ Beta$(1, 1)$, $\rho \sim$ Beta$(1.5, 1.5)$, $\rho \sim$ Beta$(2.5, 2.5)$ and $\rho \sim$ Beta$(3, 3)$ are considered, the Bayesian median widths for all parameters are always less than the corresponding widths for the likelihood estimates.

The gamma priors for $1/\sigma_b^2$ are not suitable for estimating either fixed effects or variance components. They produce severely biased estimates for the variance component and high coverage probabilities for all three parameters, particularly when sample size is small as here, for instance, with $n = 6$. Perhaps, the biased estimates along with high coverage emanate from the sensitive behaviour of gamma priors. Actually, the gamma or inverse gamma forms with small parameter values are not uninformative in any sense and can produce substantive sensitivity into the posterior specification (Gill, 2014; Gelman et al., 2006; Hodges and Sargent, 2001; Natarajan and Kass, 2000). Coverage probabilities are acceptable for fixed effects using the priors Beta(1, 1) to Beta(3, 3) for ρ as these are close to the nominal coverage. However for the variance component σ_b^2, there appeared to be overcoverage which might lead to erroneous conclusions about σ_b^2.

In Table 10, the number of tests has been increased from 6 to 40 to make comparisons, again assuming two tests per day. The first 30 tests (days 1-15) were assumed to be on base fuel B and the last 10 tests (days 16-20) on test fuel T. The results are improved in terms of bias, root mean squared error (RMSE) or root median squared error (RMdSE), coverage probability, and average or median widths. It seems that both MLE and Bayesian methods, particularly with priors $\rho \sim$Beta(1, 1) or $\rho \sim$ Beta(1.5, 1.5), produce good results in terms of relative bias.

However, the Bayesian estimates, particularly with priors $\rho \sim$Beta(1, 1) to $\rho \sim$Beta(3, 3), performed better than MLE in terms of the average and median width of 95% CIs, except for the ML estimate of β_2. All of these sets provide precise fixed effect estimates. Although the variance component σ_b^2 is estimated well in the ML method, the average width of the 95% CI is infinite as, despite the increase in the number of tests from 6 to 40, confidence intervals of infinite width were still obtained for some of the simulated 2000 data sets for the reasons discussed earlier. The data sets where the SE of σ_b^2 was not available have been excluded from the average values in Table 10.

7 Kernel Density of Simulated Estimates

Kernel density estimation can be used to visualize the sampling distribution of simulated estimates. Figure 1 shows the kernel densities of the estimates from the simulated fuel economy experiments. During the generation of samples the true parameter values were $\alpha = 32$, $\beta_2 = 1.4$, $\sigma_b^2 = 0.05$, σ^2=0.05, the sample size was 40, the number of simulations was 2000 and the priors in the Bayesian analysis were $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$, and $\rho \sim$Beta(1, 1). In the figure, the shapes of kernel densities corresponding to base fuel (α) and fuel difference (β_2) are approximately normal both in likelihood and Bayesian methods. The kernel

Table 10: Simulated performance of likelihood and Bayesian estimates in fuel economy experiments. True parameter values: $\alpha = 32, \beta_2 = 1.4$, $\sigma_b^2 = 0.05$, $\sigma^2 = 0.05$; Priors: $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$, $\log(\sigma) \sim U(-20, 20)$; simulation size=2000, $n = 40$

		Mean/Posterior Mean			Median/Posterior Median					
			% Relative			% Relative		Coverage	Average width	Median width
	parameter	Bias	Bias	RMSE	Bias	Bias	RMdSE	Probability	of 95% CI	of 95% CI
MLE	α	0.001	0.00	0.087	0.002	0.01	0.057	0.947	0.359	0.356
	β_2	-0.002	-0.12	0.125	-0.001	-0.07	0.086	0.952	0.511	0.508
	σ_b^2	0.001	2.17	0.026	-0.002	-3.50	0.018	0.976	∞	0.130
$\rho \sim$ Beta(1, 1)	α	0.005	0.01	0.071	0.005	0.01	0.071	0.950	0.283	0.280
	β_2	-0.004	-0.31	0.140	-0.004	-0.32	0.140	0.942	0.566	0.557
	σ_b^2	0.002	3.86	0.026	-0.003	-6.93	0.023	0.956	0.111	0.105
$\rho \sim$ Beta(1.5, 1.5)	α	0.007	0.02	0.070	0.007	0.02	0.070	0.959	0.287	0.290
	β_2	-0.004	-0.26	0.145	-0.004	-0.30	0.145	0.939	0.574	0.571
	σ_b^2	0.004	8.58	0.023	-0.001	-1.69	0.021	0.966	0.107	0.104
$\rho \sim$ Beta(2.5, 2.5)	α	0.008	0.03	0.071	0.008	0.03	0.071	0.957	0.285	0.280
	β_2	-0.004	-0.29	0.137	-0.005	-0.34	0.137	0.945	0.569	0.563
	σ_b^2	0.004	7.54	0.020	-0.001	-1.84	0.019	0.980	0.098	0.095
$\rho \sim$ Beta(3, 3)	α	0.008	0.02	0.073	0.007	0.02	0.073	0.948	0.287	0.290
	β_2	-0.003	-0.25	0.139	-0.004	-0.28	0.139	0.963	0.573	0.570
	σ_b^2	0.005	9.21	0.020	-0.000	-0.08	0.018	0.985	0.097	0.094
$1/\sigma_b^2 \sim$ Gamma(0.01, 0.001)	α	0.004	0.01	0.071	0.005	0.02	0.071	0.983	0.323	0.320
	β_2	-0.002	-0.17	0.142	-0.005	-0.36	0.142	0.969	0.615	0.610
	σ_b^2	-0.035	-70.97	0.036	-0.043	-85.43	0.043	0.669	0.066	0.059
$1/\sigma_b^2 \sim$ Gamma(0.1, 0.1)	α	0.014	0.04	0.073	0.015	0.05	0.073	0.992	0.376	0.370
	β_2	0.001	0.09	0.141	0.002	0.14	0.141	0.989	0.733	0.729
	σ_b^2	0.016	32.19	0.021	0.008	16.22	0.014	1.000	0.138	0.133

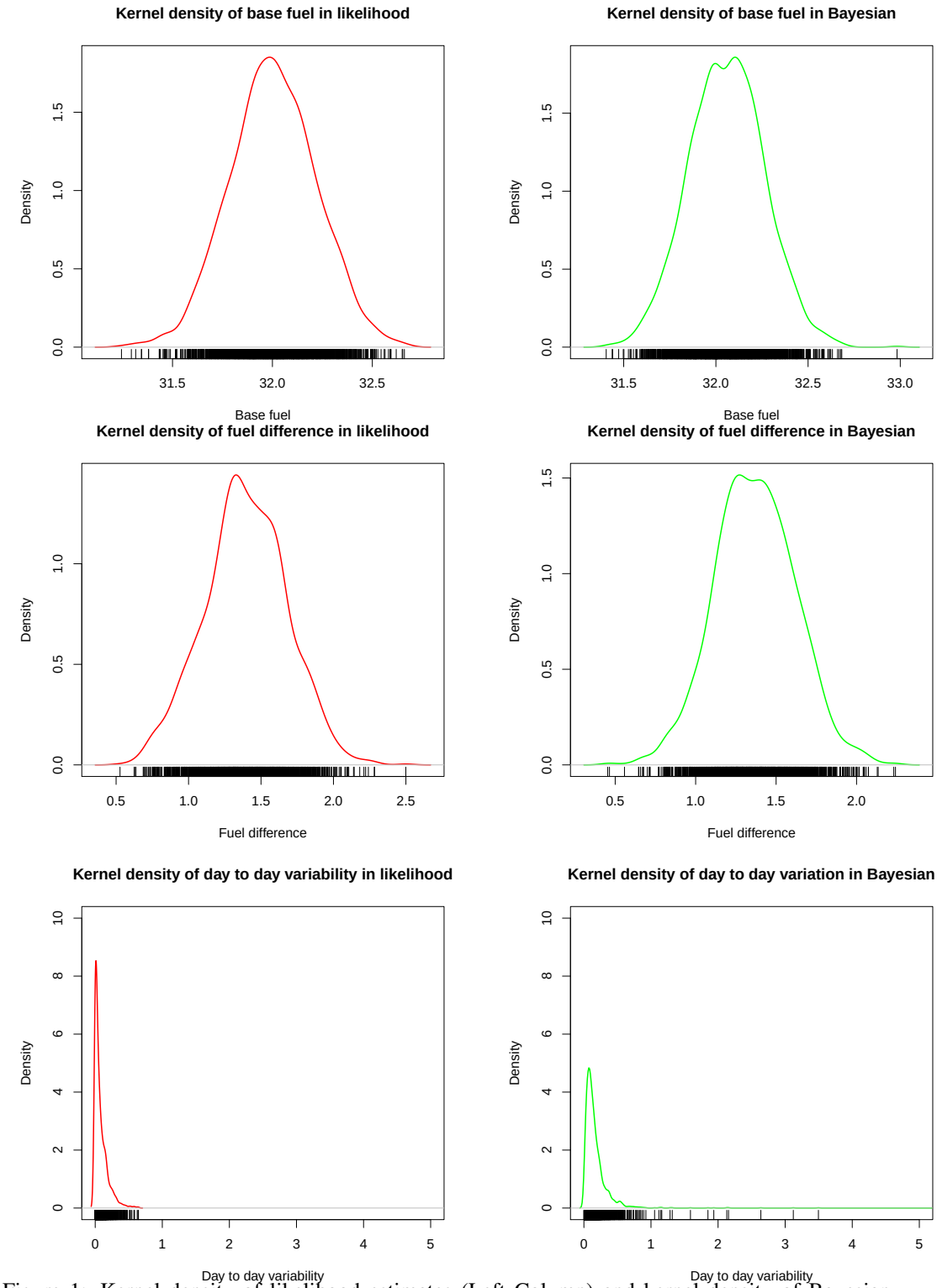

Figure 1: Kernel density of likelihood estimates (Left Column) and kernel density of Bayesian estimates (Right Column)

densities indicate that the distributions of day to day variability (σ_b^2) under likelihood and Bayesian methods are positively skewed. There were more zero estimates for day-to-day variability for the likelihood method than the Bayesian method. Therefore it is likely that zero estimates of day to day variation will be commonplace in real fuel economy experiments.

To have a better sense of why the density plots have higher bumps at certain places, we look at rug plots. A rug plot is a plot of tick marks along the horizontal axis indicating where the data are located. Clearly, there are more data in the neighbourhood between 31.5 and 32.5 where highest 'bump' is located for base fuel (α) both in the likelihood and Bayesian simulation studies. The kernel densities and rug plots for the fuel difference (β_2) also have the same pattern in both methods. However, day to day variability (σ_b^2) has different distributional patterns for likelihood and Bayesian estimates.

Based on the plots in Figure 1, it would seem that the likelihood and Bayesian methods performed fairly similarly. Nevertheless we were only able to successfully obtain likelihood estimates in 1638 of the 2000 simulations. The fitting procedure in R failed for the other 362. The Bayesian method did not encounter such problems. This is the main advantage of using Bayesian methods in fuel economy experiments. Also, in terms of coverage probabilities and width of 95% confidence/credible intervals Bayesian methods are better than likelihood based methods as they provide good coverage estimates which are close to the 95% nominal confidence limit and have less width in comparison to likelihood method (see Table 10).

8 Conclusion

In this paper, we have applied Bayesian methods to the analysis of small fuel economy experiments, which is a novel approach in this field. This study has enabled estimates of variance components to be obtained which were poorly estimable in classical methods due to the small number of groups. Likelihood-based REML and ML methods estimated the variance component due to days to be zero in the data set studied, which was considered to be unrealistic. We have implemented Bayesian techniques assuming some priors to determine this day-to-day variance component.

As the standard asymptotic theory breaks down in the case of deriving confidence intervals for the day-to-day variance component in the likelihood method, we have compared the Bayesian 95% credible interval for this variance component with the confidence intervals based on profile likelihood and bootstrap methods. However, the Bayesian estimate of the variance component is inflated as evidenced by Lambert et al. (2005) who noted that MCMC methods provide estimates of variance parameters that are upwardly biased. To evaluate the quality of Bayesian as well as likelihood estimates, we have performed simulation studies. In this regard, the frequentist properties of bias, accuracy, and coverage of the parameter estimates have been investigated.

The analysis of real fuel economy data shows that the fixed effect estimates are similar both in the classical and Bayesian methods. However, the estimates of variance components differ substantially. For instance, day-to-day variation in the model corresponding to contrast T-B was close to zero in the classical method, whereas the Bayesian estimate was 0.059 (see Table 2, Table 5). Though Bayesian methods ensure that the variance component is not estimated to be zero, Bayesian estimation is not free from criticism as it suffers from overestimation of the point estimates. Perhaps, the problem of overestimation in variance components could be reduced by assuming conservative priors for the variance parameters (Gustafson et al., 2006). However, the Bayesian estimate of the variance component has been compared with profile likelihood and bootstrap based

intervals. This showed that the Bayesian estimate was not unreasonable as the Bayesian point estimate of the day to day variance was 0.059 which lies within the profile likelihood based interval (0.000, 0.099) shown in Table 8.

In the small fuel economy experiment simulation with n=6 in Table 9, it seems that fixed effect estimates are reasonable in terms of bias, coverage probability and width, both in likelihood and Bayesian methods. However, the mean width of 95% confidence intervals is infinity in the former case as at least one width in the simulated samples is infinity due to the upper limit of that interval being infinity. Therefore, a mean based estimate of the variance component σ_b^2 is not acceptable. With respect to median based estimates, a Bayesian approach performed better than the likelihood method, particularly with priors with $\rho \sim$Beta(1, 1) and $\rho \sim$Beta(1.5, 1.5).

For fixed effects the main concern is to estimate the fuel difference β_2. Though likelihood and Bayesian estimates of β_2 are fairly close, the corresponding coverage probability in the likelihood method is slightly below the nominal level 0.95. However, for small samples, it seems that the likelihood method underestimates the variance component σ_b^2 while the Bayesian method overestimates it. From our intuition we conclude that none of the estimates of σ_b^2 obtained by the likelihood or Bayesian methods is accurate, rather perhaps it lies between the likelihood and Bayesian estimates. The mean or median widths of the Bayesian credible intervals are smaller than the median widths of the classical confidence intervals, particularly with the set of priors with $\rho \sim$Beta(1, 1) to $\rho \sim$Beta(3, 3). Among the priors the set with $\rho \sim$Beta(3, 3) performs best for small sample fuel economy experiments. When the sample size is increased from 6 to 40, the estimates are improved by providing less bias, close to desired coverage probabilities, and smaller widths of intervals (see Table 10). However, the likelihood and Bayesian methods perform fairly close to each other except for the pitfall in the average width of the 95% confidence interval of σ_b^2 in the likelihood method.

In summary, the newly applied Bayesian methods offer an alternative method of analysis of the fuel economy data that does much to address the problems of estimating variance components from small sample sizes. For Shell Global Solutions, it can be used in conjunction with traditional classical methods of analysis to further underpin the robustness of conclusions. These techniques of analyzing fuel economy have wider applicability and could be replicated in other small scale industrial experiments.

Acknowledgments

The first author acknowledges Queen Mary University of London for supporting this fuel economy research through the ImpactQM Knowledge Transfer Project and also he is thankful to Shell Global Solutions (UK) for a placement and generous funding for travel and local hospitality.

Appendix

Table A1: Data before averaging over back-to-back tests

Week	Day	Session	Treatment	Y	Week	Day	Session	Treatment	Y
1	1	am	B	31.90993	2	4	am	B	32.11118
1	1	am	B	31.61670	2	4	am	B	32.41172
1	1	am	B	32.07328	2	4	am	B	32.43854
1	1	pm	B	32.38294	2	4	pm	B	32.08281
1	1	pm	B	32.35951	2	4	pm	B	32.60450
1	1	pm	B	32.51994	2	4	pm	B	32.17017
1	2	am	B	31.92975	2	5	am	B	32.08908
1	2	am	B	32.32851	2	5	am	B	32.15086
1	2	am	B	31.67399	2	5	am	B	31.69741
1	2	pm	B	31.79294	2	5	pm	T	33.87101
1	2	pm	B	31.49287	2	5	pm	T	33.24747
1	2	pm	B	31.44593	2	5	pm	T	33.71225
1	3	am	B	31.76795	2	6	am	T	33.08393
1	3	am	B	31.36462	2	6	am	T	33.62343
1	3	am	B	31.82962	2	6	am	T	33.31008
1	3	pm	B	31.87879	2	6	pm	T	33.86173
1	3	pm	B	32.04046	2	6	pm	T	33.89393
1	3	pm	B	31.89035	2	6	pm	T	33.14460

Table A2: Data averaged over back-to-back repeats

Week	Day	Session	Treatment	Y	Week	Day	Session	Treatment	Y
1	1	am	B	31.86663	2	1	am	B	32.32048
1	1	pm	B	32.42080	2	1	pm	B	32.28583
1	2	am	B	31.97741	2	2	am	B	31.97912
1	2	pm	B	31.57725	2	2	pm	T	33.61024
1	3	am	B	31.65406	2	3	am	T	33.33915
1	3	pm	B	31.93653	2	3	pm	T	33.63342

Table A3: Data to test contrast T-B

Week	Day	Session	Treatment	Y
2	1	am	B	32.32048
2	1	pm	B	32.28583
2	2	am	B	31.97912
2	2	pm	T	33.61024
2	3	am	T	33.33915
2	3	pm	T	33.63342

Table A4: Data for contrast (T-B)-(B2-B1)

Day	Fuel	Y	Day	Fuel	Y	Day	Fuel	Y
1	B1	31.8666	3	B2	31.6541	5	B	31.9791
1	B1	32.4208	3	B2	31.9365	5	T	33.6102
2	B1	31.9774	4	B	32.3205	6	T	33.3391
2	B2	31.5772	4	B	32.2858	6	T	33.6334

References

Congdon, P. (2007), *Bayesian Statistical Modelling*, John Wiley & Sons.

Galindo-Garre, F., Vermunt, J. K., and Bergsma, W. P. (2004), "Bayesian posterior estimation of logit parameters with small samples," *Sociological Methods & Research*, 33, 88–117.

Gelman, A. et al. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–534.

Gill, J. (2014), *Bayesian Methods: A Social and Behavioral Sciences Approach*, CRC press.

Gilmour, S. G. and Goos, P. (2009), "Analysis of data from nonorthogonal multistratum designs in industrial experiments," *Applied Statistics*, 58, 467–484.

Gustafson, P., Hossain, S., and Macnab, Y. C. (2006), "Conservative prior distributions for variance parameters in hierarchical models," *Canadian Journal of Statistics*, 34, 377–390.

Hodges, J. S. and Sargent, D. J. (2001), "Counting degrees of freedom in hierarchical and other richly-parameterised models," *Biometrika*, 88, 367–379.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and Jones, D. R. (2005), "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS," *Statistics in Medicine*, 24, 2401–2428.

Litière, S., Alonso, A., and Molenberghs, G. (2008), "The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models," *Statistics in Medicine*, 27, 3125–3144.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, 10, 325–337.

Natarajan, R. and Kass, R. E. (2000), "Reference Bayesian methods for generalized linear mixed models," *Journal of the American Statistical Association*, 95, 227–237.

Pinheiro, J. C. and Bates, D. M. (2000), *Linear Mixed-Effects Models: Basic Concepts and Examples*, Springer.

Rahman, M. L. (2015), "Bayesian Analysis of Multi-Stratum Designs and Probability-based Optimal Designs with Separation." Ph.D. thesis, Queen Mary University of London.

Received: *March 29, 2020*

Accepted: *May 14, 2020*