Final Assignment: Predicting Director Compensation

Jasper Ginn (s6100848)

April 8, 2019

Introduction

Problem statement, literature & hypotheses

Methods & Results

Table XX below shows the descriptive statistics for each variable. The outcome variable **compensation** is given in thousands of Great British pounds and has been log-transformed.

Model 1 is given by the following equation:

$$\log(\widehat{\text{compensation}_i}) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{male}_i + \epsilon_i \tag{1}$$

Model 2 is given by the following equation:

$$\log(\widehat{\text{compensation}_i}) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{male}_i + \beta_3 \text{SectorServices}_i + \beta_4 \text{SectorBasicMaterials}_i + \epsilon_i$$

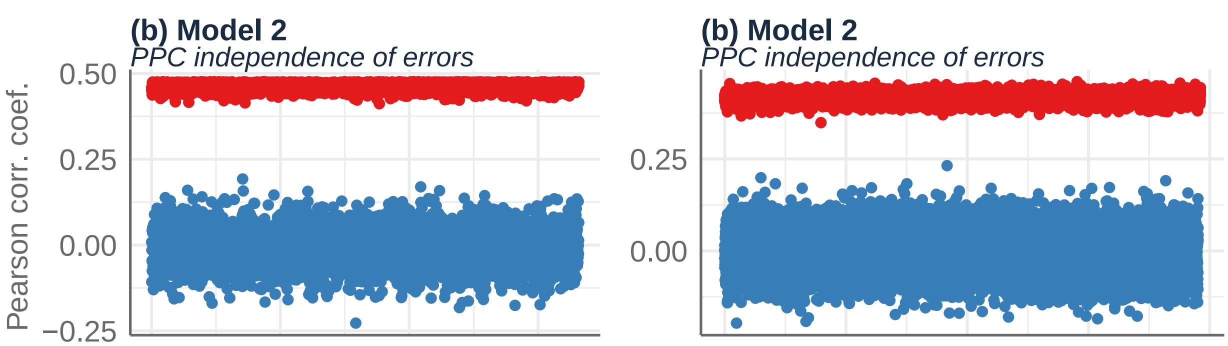

Figure 1: A random subset of observed and simulated correlation coefficients for model 1 (left) and model 2 (right). The figure shows that the observed residuals (in red) are much more correlated than is reasonable under the model (represented by the blue, simulated values).

From figure XX and the posterior predictive checks (table XY), we observe that the data are not independent ($\hat{r}_{\text{PPC, Model1}} = .466$). That is, directors who are in the same board tend to are more similar to each other than directors from other boards. This is not unexpected: the data are hierarchical in nature such that individuals are nested in boards. Indeed, the results of the random effects model shows that the intra-class correlation coefficient ρ equals 57%, meaning that the expected correlation of two randomly picked directors from the same company is $\hat{r} = .573$.[1]

The final random effects model is presented in table YY. This model corresponds to the following equation.

[1] The results of running all different stages of a multilevel model are presented in another document.

$$\text{compensation}_{ij} = \gamma_{00} + u_{0j} + \gamma_{10} \cdot \text{Age}_{ij} + e_{ij} \tag{3}$$

Where γ_{00} is the overall intercept and u_{0j} is a company-specific error term. Notice that the fit of this model is much better than that of the previous models, indicating that the multilevel approach seems appropriate. The marginal and conditional R-squared values [CITE] are $R^2_{\mathrm{M}} = .0056;\ 95\%\ \mathrm{CCI} = [.0008, .015]$ and $R^2_{\mathrm{C}} = .5398;\ 95\%\ \mathrm{CCI} = [.4807, .6]$ respectively. This indicates that the fixed part of the model (age) explains almost no variation in the data, but the fixed and random parts together explain some 54% of the total variation. Hence, we must conclude that we do not have the right variables at either level 1 or 2 that would help us explain the heterogeneity in comepnsation among directors.

	Dependent variable:		
	Compensation (GBR '000, logged)		
	(1)	(2)	(3)
	Linear	Linear	Linear mixed effects
	(blm)	(blm)	(JAGS)
(a) Fixed			
Constant	4.991 (4.936, 5.046)	4.821 (4.732, 4.910)	4.991 (4.877, 5.110)
SectorBasic Materials		.225 (.079, .370)	
SectorServices		.292 (.169, .414)	
Male	.065 (−.085, .215)	.153 (.095, .209)	
Age	.008 (0.000, .017)	.009 (.001, .017)	.001 (.0035, .0155)
(b) Random			
σ_e^2	.516	.501	.342
σ_{u0}^2			.399
(c) Model Fit			
Observations	336	336	336
Companies			52
DIC	508	488	280
Penalty.	4	6	47
R^2	.021	.098	.006 (M), .536 (C)
BF	.112	37.596	
(d) Post. Pred. Checks			
Normality	.361	.353	
Homoskedasticity	.326	.607	
Independence	0	0	
Note:		Baseline is sector 'Financials' for models (1) and (2)	

Table 1: Model results

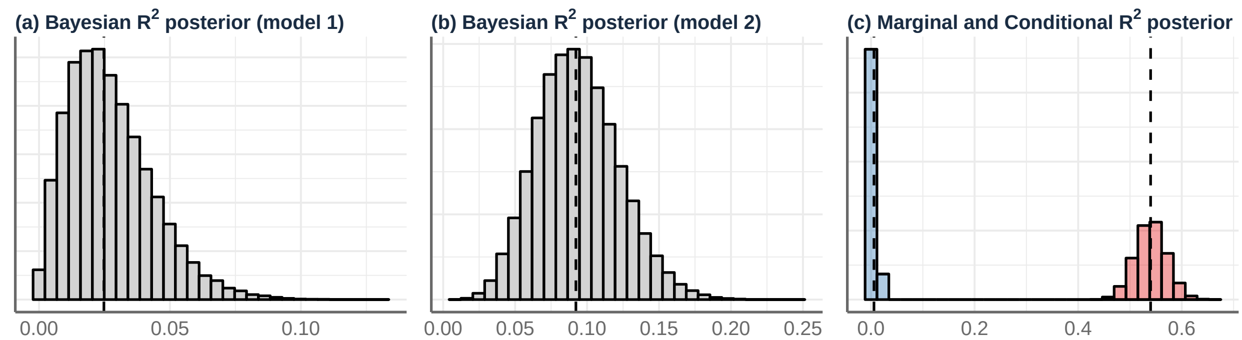

Figure 2: Bayesian R-squared value for model 1 (a) and model 2 (b). The proportion of cases in which the R-squared value of model 2 exceeds that of model 1 is .99. Figure (c) shows the marginal (blue) and conditional (red) R-squared values for the posterior distribution. The marginal R-squared indicates the amount of variance explained by the fixed part of the model; the conditional R-squared indicates the amount of variance explained by the fixed and random part of the model.

Differences in Bayesian and Frequentist inference

- Interpretation of random variables
- Same rules for probability, but different approach

–> Probability of the null hypothesis (p 161, 162 Berger & Benny). "Flipping" of the hypothesis

- Can answer the same questions, but Bayesian more flexible
- Endless transformations on posterior distribution
- Frequentists obscure the interpretation process, Bayesian obscure the posterior distribution collection process

Evidence of the the idea that we (humans) are not to be trusted when it comes to the interpretation of numbers abounds in the literature of behavioral economics. [CITE KAHNEMAN].

Frequentist statistics obscures the interpretation of critical statistics to the point where students learn the heuristics ('p-value is significant') before truly understanding what that heuristic means. Conversely, Bayesian statistics provides an intuitive interpretation of statistical results but obfuscates the estimation method by using the arcane process of Markov Chain Monte Carlo (MCMC) sampling.

- Incorporating domain knowledge

In frequentist statistical inference, evidence may originate only from the data. This is not the case in Bayesian statistical inference, where inferences are based on a mix of domain expertise ($prior/belief$) and evidence from the data.

The claim that the Frequentist approach is more objective would hold only in a universe where the data collection process is guaranteed to be objective. This is a tenuous assumption at best in social science data (how else would we get by that memorable phrase 'lies, damned lies and statistics'?); the way in which we collect data is fraught with subjective decisions during data collection, manipulation and analysis [CITE]. Hence, many authors have argued that data, in and of itself, is not objective [CITE], and the critique leveled at Bayesians boils down to the practice of incorporating domain knowledge *explicitly* through the use of a prior. It does, however, raise the question of the extent to which the use of priors influence the analysis.

This is illustrated in figure XX below in the case of the directors data. Here, we show the result of increasing the certainty of our prior belief (translates to smaller variance).

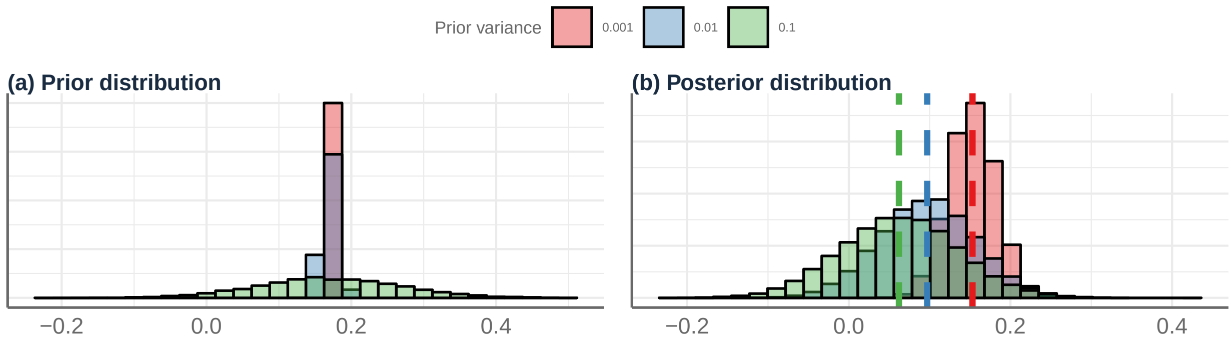

Figure 3: Effect of adjusting the prior variance of the coefficient for variable Male (a) on the posterior distribution (b). The mean is set at M=.17 which represents a 17 percent increase in compensation for males versus females. Reducing the prior variance represents increased certainty about the estimate of our domain knowledge and weights the information from the data less strongly compared to domain knowledge. The dashed lines indicate the value of the posterior means.

The effect of changing the prior variance can be summarized using a posterior shrinking factor (CITE BETANCOURT) and posterior z-score. The shrinking factor shows us the factor by which the variance of the posterior distribution shrinks or expands compared to the prior variance. The posterior z-score tells us the direction and magnitude by which the posterior mean shifts compared to the prior mean; if we are confident in the precision of our domain knowledge (resulting in small prior variance), the resulting posterior weights this information strongly and we end up with a posterior mean that lies closer to the prior mean (represented by the z-score). However, given that we have small variance, and if this is not corroborated by the data, then the shrinkage factor will be large.

If we carry this argument to its logical endpoint, it follows that we would next arrive at a means to compare these posteriors relative to each other. This, in essence, is the Bayes Factor.

Other material

If we desire to interpret this p-value like we do the traditional, Fisherian p-value, then we must assume that, under the assumption that a particular test is not violated, the Bayesian p-value is uniformly distributed. Otherwise, we cannot interpret it as a proportion. Hence, we desire $p_{\text{posterior}} = P(x \leq X)$. We often find that this assumption does not hold true for posterior predictive checks. To this end, we can simulate the assumption when it holds [CHANGE]. This is illustrated in figure XX for the posterior predictive checks included in the R library blm.

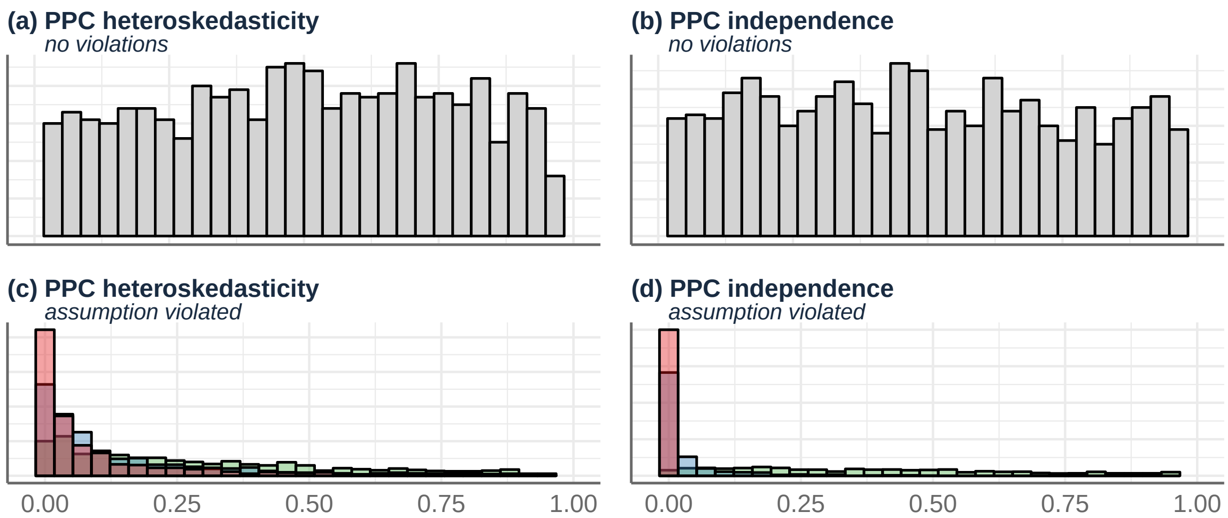

Figure 4: Distributions of posterior predictive p-values for 1.000 simulated data sets. In plots (a) and (b), the simulated data are drawn from a normal without any violations of the linear regression assumptions. In plot (c), the assumption of homoskedasticity is violated in each of the simulations. In plot (d), the assumption of independence of errors is violated in each of the simulations. The color indicates the severity of the violation; the green bars indicate mild violation, blue indicates medium violation and red indicates severe violation. The script used to generate the data and run the simulations may be found here.

References

Lynch, S. M. (2007). Introduction to applied Bayesian statistics and estimation for social scientists. Springer Science & Business Media. Chapter 9.2

Aarts, E. (2019). Introduction to multilevel analysis and the basic two-level regression model, week 1 notes [powerpoint presentation]. *Introduction to Multilevel Analysis*, Utrecht University.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution, 4(2), 133-142.