The Gender Gap in Independent Director Compensation

Jasper Ginn (s6100848)

April 8, 2019

Introduction

The gender gap in compensation remains a pervasive problem despite efforts to achieve income parity between men and women. In this paper, I examine the gender gap in compensation among 336 independent directors in three sectors and 52 companies using Bayesian Linear Regression. The boards of American companies consists of *executive* and *independent* directors. Execute directors are employed by the firm and usually have high managerial positions within the company, while independent directors are concerned primarily with oversight.

The results indicate that we cannot conclude that male independent directors receive a higher remuneration than female independent directors after controlling for age. Model fit statistics and evaluation suggest that linear regression is not an appropriate method to analyse these data because director compensation is highly correlated within companies.

The paper is structured as follows. Section one elaborates on the differences between Frequentist and Bayesian statistical inference. In turn, section two outlines the hypotheses, model specification and reports the results of the analyses. Section three concludes.

Differences in Bayesian and Frequentist Inference

Many differences between Frequentist and Bayesian inference originate in their respective view of uncertainty, how this is captured by probabilities and what causes this uncertainty. In the Frequentist framework, a probability is defined as the number of times an event will occur in the long run. In other words, it is the limiting value of m successes in a sequence of n trials for a particular event, or $p = \lim_{n \to \infty} \frac{m}{n}$. (I. Miller, Miller, and Freund 2014, 21). Hence, talking about probability makes sense only in the context of infinite trials and probabilities converge to some fixed quantity as the number of trials go to infinity. Our uncertainty about the true value of the probability reduces as we repeatedly take more (or larger) samples. In turn, this implies that the only source of randomness by which our estimate $\hat{p}$ differs from the true value p comes from the data, which may differ from sample to sample due to, for example, sampling error. (I. Miller, Miller, and Freund 2014)

The Bayesian framework considers probabilities as a means of quantifying uncertainty about knowledge (Gelman et al. 2013, 11–13). Even though the 'true' parameter value may be fixed, we are limited by our knowledge of this value. The uncertainty in our knowledge will hopefully (but not necessarily) decrease as we collect more information and is represented by the posterior distribution.

Objective and Subjective Knowledge

A Frequentist believes that evidence can only originate from the data. This is not the case in Bayesian statistical inference, where inferences are based on a mix of domain expertise (*prior* or *belief*) and evidence from the data. This makes perfect sense in the Bayesian framework; if we are certain about our knowledge then we can constrain the parameter space by injecting what we know *a priori*. In other words, a Bayesian looks upon prior beliefs as just another source of knowledge that have been translated into a probability density.

The claim that the Frequentist approach is more objective only holds if one regards the data collection process, (pre-)processing steps and analysis as guaranteed to be objective. This is a tenuous assumption at best in the social sciences; statistics is fraught with subjective decisions during data collection, manipulation and analysis (Berger and Berry 1988), and the charge of subjectivity leveled at Bayesians boils down to the practice of incorporating domain knowledge *explicitly* through the use of a prior distribution.

Methods of Estimation and Hypothesis Testing

Whether a statistician regards the data or the parameters as a random variable determines their choice of estimation method. A Frequentist will want to find the most likely combination of parameters $\hat{\theta}$ that explain the data and that provide consistent and asymptotically unbiased estimators. Conversely, given that a Bayesian thinks of the data as fixed and the parameters as random variables, they are interested in finding the distribution of the parameters and hence the source of uncertainty in our beliefs after seeing the evidence (I. Miller, Miller, and Freund 2014, ch.8).

The way in which we think about statistical inference under these frameworks is illustrated by the difference in interpretation of the confidence interval and the credible interval. When we calculate the confidence interval, the upper and lower confidence limited should be interpreted as random variables because they vary across samples. Crucially, they do not reflect probabilities but confidence in our point estimates (I. Miller, Miller, and Freund 2014, 317–20). A credible interval, being associated with a posterior distribution, necessarily deals with probability statements such that we can say that the parameter is contained in the credible interval with some probability (e.g. 95%) (Lynch 2007, 58).

Further implications are to be found in the way we test hypotheses in these frameworks. In the Frequentist framework, we usually partition the parameter space into an acceptance and a rejection region based on some null and alternative hypothesis. On the basis of a test statistic, computed from the data, we then decide whether or not the result we observe is likely to occur due to chance. We use either confidence intervals or p-values to express support for the alternative hypothesis, but, importantly, we can never quantify support in favor of the null hypothesis (I. Miller, Miller, and Freund 2014, 337–39). This is not so in the Bayesian Framework. Given that the data can be viewed as a means to update a prior belief, hypothesis testing using Bayes Factors is a relative statement about the degree to which the evidence found in the data supports one hypothesis over another (Hoijtink et al. 2019).

Methods & Results

This section describes the results of the statistical analysis. The outcome variable compensation is given in thousands of Great British Pounds and has been log-transformed. The variable age has been grand-mean centered to facilitate interpretation and estimation. Some 44% of the data has been deleted due to missingness on one of the variables, which leaves us with 336 observations, 60 of whom are female and 276 are male. The model is given by the following linear regression equation:

$$\log(\widehat{\text{compensation}}_i) = \beta_0 + \beta_1[\text{age}_i - \overline{\text{age}}] + \beta_2\text{male}_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_e^2)$$

Given that we measure the compensation on a log scale, the coefficients we derive from the model must be interpreted as change in percentages. This forces us to rethink the priors we specify. We assume that, as age increases, compensation increases as well. Given that we are not certain about the extent to which it increases, we set the the mean to .05 and the standard deviation of this estimate to .2, representing a spread of approximately 20%. We know from earlier studies that male executive directors earn $8 - 25\%$ more than female directors (Bell 2005). Due to the lack of research with respect to the remuneration of independent directors, we will take this as a baseline and set this prior a mean of .165 with a variance of .1, which reflects our uncertainty of the prior knowledge.

In accordance with our expectations, we hypothesize that male directors earn some $5 - 20\%$ more than their female counterparts (H_1) and that compensation only increases with age (H_2). H_u is the unconstrained hypothesis.

$$\mathrm{H}_1\text{: } .05 < \beta_{\text{Male}} < .20 \qquad \mathrm{H}_2\text{: } \beta_{\text{Age}} > 0 \qquad \mathrm{H}_u\text{: } \beta_{\text{Male}}, \beta_{\text{Age}}$$

The model results are presented in table 1. The variable age is sampled using a random-walk Metropolis-Hastings (MH) step. Unlike the Gibbs sampler, which samples full conditional posteriors, the MH step draws samples from a non-normalized posterior distribution using a normal proposal density. The convergence diagnostics show that the posterior distributions have forgotten their initial states and have converged to their stable distributions. There is no trace of autocorrelation and the Gelman-Rubin statistic is close to 1, indicating that the chains have converged to the same stationary distribution. Roughly 43% of samples for age drawn using the MH algorithm are accepted. The DIC for the model of interest (model 2) is 509 compared to 513 for the intercept-only model (model 1). The difference between the DIC scores indicates that the model fits better than the intercept-only model, but the difference in scores of 4 does not convince that model 2 is an improvement over model 1. The MC error, which should be less than 5% of the standard deviation, is negligible.

The model results indicate that gender is not a predictor of director compensation ($\beta_{\text{male}} = .071; 95\%\ \text{CCI} = [-.075, .219]$). The CCI and standard deviation for this coefficient indicate that there is a lot of uncertainty in this estimate. Age is also not a predictor and has a small, positive effect on director compensation ($\beta_{\text{age}} = .008; 95\%\ \text{CCI} = [-.00, .017]$). The median R^2 value for this model is $R^2 = .021; 95\%\ \text{CCI} = [.002, .059]$) (Gelman et al. 2018), indicating that we explain some 2% of the variance in compensation.

The support for in the data H_1 versus H_c[1] is very small ($BF = 1.243$). The evidence that age has a positive impact on compensation (H_2) has more support after observing the data ($BF = 24.9$), and H_2 has 20 times more support in the data than H_1. The posterior model probabilities (PMPb in table 1) indicate that we should prefer H_2 over H_1. However, this preference comes with a high error probability (approximately 55%).

Figure 1: 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 posterior predictive checks (table 1) suggest that the data violate the independence assumption ($p_{\text{independence}} = 0$), which is checked by repeatedly simulating outcome data under the model and comparing

[1] H_c is the complement of H_1. Given that we only have inequality constraints, H_1 is evaluated against H_c and not H_u. See (Hoijtink et al. 2019)

the correlation of lagged residuals on simulated and observed outcome data. The posterior predictive p-value then represents the proportion of cases where $r_{\text{sim}} > r_{\text{obs}}$. Both of the posterior predictive checks included in the R library blm show an approximately uniform distribution when there are no violations (figure 1).

This is not unexpected: the data are hierarchical in nature such that individuals are nested in boards and directors tend to are more similar to their co-directors in the same boards in terms of compensation. Given the hierarchical nature of the data, we next run a linear mixed effects model. The intra-class correlation coefficient shows that a large portion of the variance is found at the company level ($\rho = 57\%;\ 95\% CCI = [.451, .686]$), meaning that the expected correlation of two randomly picked directors from the same company is $\hat{r} = .573$.[2] The final model corresponds to the following equation.

$$\log(\widehat{\text{compensation}}_i) = \gamma_{00} + u_{0j} + \gamma_{10} \cdot [\text{Age}_{ij} - \overline{\text{Age}}]) + e_{ij}$$

Where γ_{00} is the overall intercept and u_{0j} is a company-specific error term. The DIC of 281 indicates that the multilevel approach is appropriate, given that this model fits better than the regular regression model. The marginal and conditional R-squared values (Nakagawa and Schielzeth 2013)[3] are $R^2_{\text{M}} = .0056;\ 95\%\ \text{CCI} = [.00, .015]$ and $R^2_{\text{C}} = .5398;\ 95\%\ \text{CCI} = [.4807, .6]$ respectively. These values suggest 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.

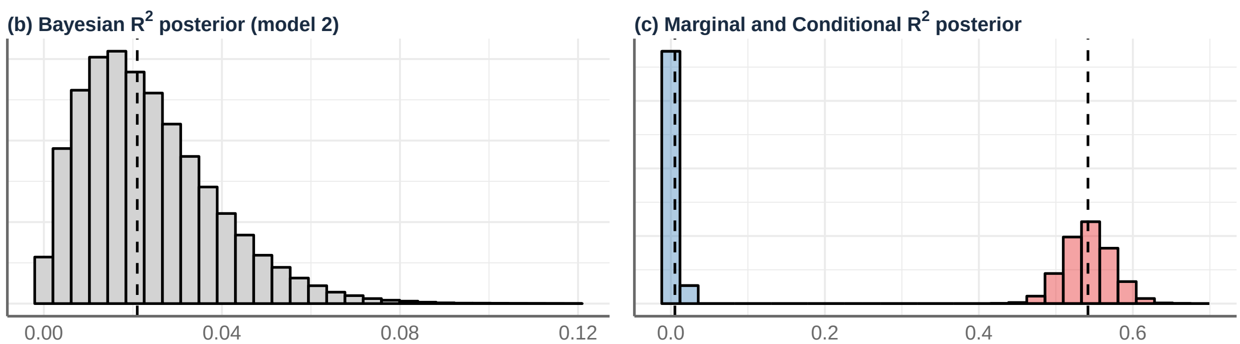

Figure 2: Bayesian R-squared value for model 1 (a) and model 2 (b). The proportion of cases in which the R-squared value of model 2 exceeds that of model 1 is .99. Figure (c) shows the marginal (blue) and conditional (red) R-squared values for the posterior distribution. The marginal R-squared indicates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

Conclusion

The analysis shows that gender is not a predictor of compensation for independent directors. There are some clear limitations to this analysis. Firstly, I have to delete some 44% of the data during missingness. Secondly, the results indicate that there is a lot of uncertainty in the estimation of the gender effect, which is exemplified by the large standard deviation of the posterior distribution. Furthermore, director compensation is best modeled using company-specific intercepts, but we do not have the right variables at either the company or individual level that would help us explain the heterogeneity in compensation among directors within boards and between boards. Future research should focus on collecting more complete data and more informative covariates.

[2] The complete multilevel analysis of this data can be found here.

[3] The marginal R-squared takes into account only the fixed part of the variance, while the conditional R-squared takes into account the fixed and random parts. See (Nakagawa and Schielzeth 2013)

	Dependent variable:		
	Compensation (GBR ’000, logged)		
	(1)	(2)	(3)
	Intercept-only	Full model	Linear mixed effects
	(blm)	(blm)	(JAGS)
(a) Fixed			
Constant	4.991 (4.935, 5.046)	4.991 (4.936, 5.046)	4.993 (4.875, 5.112)
Male		.071 (−.075, .219)	
Age		.008 (−.00, .017)	.009 (.003, .015)
(b) Random			
σ_e^2	.52 (.482, .56)	.516 (.479, .558)	.342 (.315, .372)
σ_{u0}^2			.399 (.322, .504)
(c) Model Fit			
Observations	336	336	336
Companies			52
DIC	513	509	281
Penalty.	2	4	49
R^2		.021	.006 (M), .536 (C)
(d) Post. Pred. Checks			
Normality		.362	
Homoskedasticity		.321	
Independence		0	
(e) Bayes’ Factors (model 2 only)			
Hypothesis	BF (complexity, fit)	PMPa	PMPb
$H_1 : .2 > \beta_{\text{Male}} > .05$	1.243 (.512, .566)	.403	.295
$H_2 : \beta_{\text{Age}} > 0$	24.86 (.594, .972)	.597	.438
$H_u : \beta_{\text{Male}}, \beta_{\text{Age}}$			.267

Table 1: Model results

References

Bell, Linda A. 2005. “Women-Led Firms and the Gender Gap in Top Executive Jobs.”

Berger, James O, and Donald A Berry. 1988. “Statistical Analysis and the Illusion of Objectivity.” *American Scientist* 76 (2). JSTOR: 159–65.

Gelman, Andrew, Ben Goodrich, Jonah Gabry, and Aki Vehtari. 2018. “R-Squared for Bayesian Regression Models.” *The American Statistician*, no. just-accepted. Taylor & Francis: 1–6.

Gelman, Andrew, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis*. Chapman; Hall/CRC.

Hoijtink, Herbert, Joris Mulder, Caspar J Van Lissa, and Xin Gu. 2019. “A Tutoial on Testing Hypotheses Using the Bayes Factor.” PsyArXiv.

Lynch, Scott M. 2007. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer Science & Business Media.

Miller, Irwin, Marylees Miller, and John E Freund. 2014. *John E. Freund’s Mathematical Statistics with Applications*. Boston: Pearson,

Nakagawa, Shinichi, and Holger Schielzeth. 2013. “A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models.” *Methods in Ecology and Evolution* 4 (2). Wiley Online Library: 133–42.