Evolution and Human Behavior 42 (2021) 593–596

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Commentary

Height shows no clear association with self-serving beliefs about wealth redistribution: A commentary on Richardson (2020)

Alex L. Jones

Department of Psychology, Swansea University, UK

Researchers interested in evolutionary approaches to understanding human behaviour can gain much from the application of open access 'big data' sets to test hypotheses. Richardson (2020), using the European Social Survey to leverage a large and diverse sample size, demonstrated that greater height is associated with more self-serving or less favourable attitudes towards government redistribution of wealth. Height also interacted with household income, such that taller individuals with greater income were especially self-serving. These relationships were consistent across many robustness checks and model parameterisations, suggesting a convincing effect, and is broadly consistent with existing theories that human height (as a proxy for formidability) and resource acquisition play a role in self-serving behaviour (Petersen, Sznycer, Sell, Cosmides, & Tooby, 2013; Sell, Tooby, & Cosmides, 2009).

However, the reanalysis of the data used by Richardson (2020) described here illustrates that there is no conclusive evidence for or against this effect, and highlights some common statistical issues present in the original analysis.

1. Improper choice of likelihood

The original analysis of Richardson (2020) utilised linear mixed models that accounted for country-level clustering of responses, regressing the predictors against a five-point Likert scale capturing attitudes towards government wealth redistribution. An important issue here is that a linear regression model will assume continuous, normally distributed errors when modelling the data, and thus make continuous predictions that do not fit the data well. While all models are wrong, it is important to fit a model that reflects the data-generating process as closely as possible – here, responses on a one-to-five Likert scale. Such scales are often modelled with standard linear regression with little loss in interpretation and fit, particularly when the number of scale items have five or more levels (Johnson & Creech, 1983), and are sometimes a recommended alternative choice depending on the number of categories and respondents use of the scale (Gelman, Hill, & Vehtari, 2020). However, recent demonstrations have shown this practice can lead to serious errors in interpretation (Bürkner & Vuorre, 2019; Kruschke, 2014; Liddell & Kruschke, 2018). Indeed, ordinal regression models have an appropriate likelihood function that is capable of modelling the underlying latent variable (here the attitude towards wealth redistribution) as a continuous factor that the Likert responses are practical realisations of. As shown below, a model with this form outperforms simple linear approaches.

2. The Null Hypothesis is always false – especially in large samples

Large samples are desirable for accurate estimation of effects and power for hypothesis testing. However, they come with a serious drawback – because the null hypothesis of zero effect is always false, particularly in social science research where measurements are noisy (McShane, Gal, Gelman, Robert, & Tackett, 2019), very small (and possibly meaningless) deviations from zero can be strongly statistically significant in large samples. This outcome is well described by Meehl (1967; 'Meehl's Paradox'), leading to situations where any theory can garner evidence with a large enough sample size. Indeed, the main effect of height in Richardson (2020) is acknowledged as a consistently small effect, around $b = -0.03$. While small effects can be valuable, when they are observed in large samples, particular care is needed in their interpretation. A more stringent approach is the use of a 'region of practical equivalence', or two-one sided tests (Kruschke, 2018), that specify a null *region* that an estimate must overcome to be taken credibly – that is, rejecting a null of precisely zero is not sufficient for testing theoretical claims (Kruschke & Liddell, 2018). This approach is used below and is absent from Richardson (2020).

3. Overfitting

Richardson (2020) arrives at a final model that demonstrates statistical significance of all predictors, with a higher-order interaction between height and household income. Examination of open data resources (osf.io/t9a4m/) indicates this model was arrived at by removing non-significant further interactions with height until an interaction with height remained. This is akin to stepwise regression with a focus on a given effect, pruning predictors until maximal significance is reached, which has been shown to drastically overfit data (Smith, 2018).

E-mail address: alex.l.jones@swansea.ac.uk.

https://doi.org/10.1016/j.evolhumbehav.2021.09.002

Available online 7 October 2021
1090-5138/© 2021 Elsevier Inc. All rights reserved.

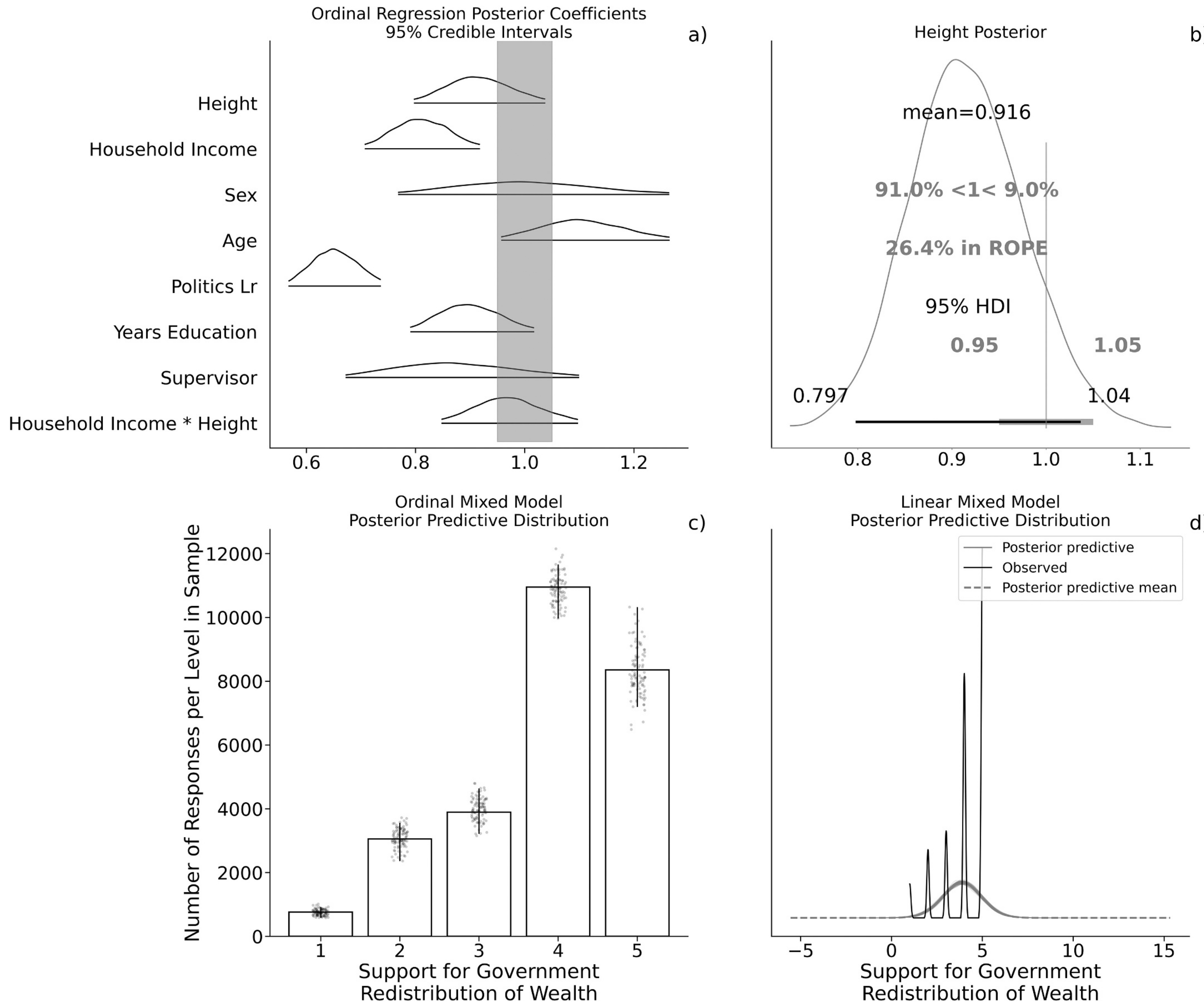

Fig. 1. Panel A shows 95% credible intervals for the posterior distributions of the coefficients. Shaded area represents the null region. Panel B shows the posterior of just the height coefficient and associated statistics. Panels C and D represent the posterior predictive distribution of the ordinal parameterisation and standard linear model – the bars in panel C represent the counts of the responses in the observed data, and the dots and error bars represent the posterior predictions and 95% credible interval of those predictions.

4. A robust Bayesian ordinal mixed model approach

The above limitations engender scepticism of the association between height, other covariates, and attitudes towards government wealth redistribution. Here, an ordinal mixed model is estimated using Bayesian methods. The final 'model A' described by Richardson (2020) is fitted, regressing scaled height, income decile, sex, age, political orientation, education, authority position at work, and the interaction between height and income onto the ordinal scale of attitudes towards government equality, for the same sample of 27,018 individuals. A random intercept was included for country. Ordinal regression coefficients represent log-proportional odds, which when exponentiated can be interpreted as the odds of a higher response being given on the ordinal scale as the predictor increases with one unit. Ordinal regression also estimates 'cutpoints' that represent the point in the latent variable where responses shift from one category to another.

To mitigate overfitting, all coefficients had a Laplace prior placed over them with mean zero and scale of 2. This is not particularly restrictive as odds of around 7 or 0.13 are still considered plausible. Random intercepts for each country were sample from a wide normal distribution with mean zero and a scale of ten. The cutpoints were sampled from a *t*-distribution with a mean of zero and a scale of 20 – this is very wide but ensures the model is less sensitive to outliers. The normality parameter for the *t*-distribution was taken from an exponential prior distribution with a rate of 1/23.

Finally, a ROPE was set for each parameter for odds between 0.95 and 1.05. That is, the posterior distributions of each coefficient would be considered null if their 95% credible interval fell entirely within this region; the null would be rejected for that coefficient if the 95% credible interval was fully outside of it, and considered a lack of conclusive evidence if it overlaps (Jones, Jaeger, & Schild, 2021).

Models were fit using automatic differentiation variational inference (ADVI; Kucukelbir, Tran, Ranganath, Gelman, & Blei, 2017) due to the size of the sample, using PyMC3 (Salvatier, Wiecki, & Fonnesbeck, 2016). Analyses can be found on the Open Science Framework (osf.io/9sxu3/).

Table 1
Posterior summaries of each coefficient, with 95% credible interval.

Predictor	Posterior Mean (in odds)	SE	Lower Credible Interval	Upper Credible Interval	Probability effect is greater than 1 (null/in opposite direction)	Probability effect is in ROPE [0.95, 1.05]
Height	0.92	1.07	0.80	1.04	0.09	0.26
Household Income	0.81	1.07	0.71	0.92	0	0.01
Sex (male)	1	1.14	0.78	1.28	0.49	0.31
Age	1.1	1.07	0.96	1.26	0.91	0.22
Politics left-right	0.65	1.07	0.57	0.74	0	0
Years Education	0.9	1.07	0.79	1.02	0.05	0.19
Authority Position at work (yes)	0.87	1.14	0.68	1.11	0.14	0.18
Height * Income	0.97	1.07	0.85	1.1	0.30	0.51

5. Results and discussion

The 95% credible intervals of the posteriors of each coefficient in the model are shown in Fig. 1a, and summarised in Table 1.

These coefficients represent the odds of providing a higher response on the ordinal scale given a one-unit increase in the predictor. Only two predictors, political orientation and household income, are sufficiently far from the null region that the null can be rejected. Notably, the effect of interest, height, shows no strong associations. For the main effect of height, the mean of the posterior is 0.92, 95% CrI [0.80. 1.04], indicating that a one standard deviation increase in height (here around 9 cm) is associated with an 8% decrease in the odds of responding with higher agreement for government wealth distribution (see Fig. 1b). This does mirror the findings of Richardson (2020), but the full posterior reveals that there is uncertainty in this estimate – indeed, there is an 9% probability the effect is actually positive, and a 26% probability the effect is within the small null region. That is, the effect may be practically null or in the opposite direction with a substantial probability. Readers are free to define their own ROPE, but excluding a positive effect is non-trivial. The same kinds of claims can be made for the height and income interaction, which shows even greater uncertainty here, with a 30% probability of being positive, and a 51% probability of the effect being with the null region, indicating no credible evidence of a clear effect..

As a further test, a posterior predictive check (Gelman & Robert, 2013) was carried out on the ordinal model. This involves randomly sampling values from the posterior and feeding them to the likelihood to generate new data. If this looks similar to the observed data, the model can be considered appropriate. For the original data, the sum of each level of the response was taken (e.g., how many responses were one). One hundred samples were drawn from the posterior predictive distribution and the same computation was undertaken. The results are plotted in Fig. 1c, showing that the model accurately recovers the overall pattern within the data. Fig. 1d shows the same pattern for the exact same model specification used by Richardson (2020), estimated using a robust Bayesian equivalent linear mixed model, demonstrating how such a model is unable to accurately predict this data.

Finally, the ordinal model here, and the Bayesian equivalent to the linear mixed model used by Richardson (2020) were compared using leave-one-out cross validation methods (Vehtari, Gelman, & Gabry, 2017). For the ordinal model, the deviance score (lower indicating a better fit) was 70,223, while for the linear model it was 76,938, a difference of 6714. In absolute terms neither model predicts the data well, but the ordinal parameterisation is far superior.

The analyses discussed here highlight the importance of modelling the data generating process as accurately as possible, using the appropriate likelihood function and model structure. While historically linear models have been utilised for their assumptions and ease of use, modern estimation procedures (particularly Bayesian methods) are capable of modelling any outcome (Kruschke & Liddell, 2018; Liddell & Kruschke, 2018).

Large scale, open-access datasets have enormous potential to inform researchers about evolutionarily relevant outcomes. The reanalyses described here show two areas of concern that researchers should pay close attention to. First, consideration should be given to specifying a region of practical equivalence (Kruschke, 2018) or a smallest effect size of interest (SESOI Lakens, Scheel, & Isager, 2018) before analysing the data, because even slight deviations from zero can be significant with enough observations. Templates for pre-registering analyses for secondary data are available via the OSF (osf.io/x4gzt). Second, the likelihood function of any analysis is a choice on the part of the researchers, and it is important to select the one that closely matches the data-generating process or check the robustness across results of different likelihoods, if the former is unclear. In this regard, Bayesian approaches offer flexible and powerful tools for modelling data and checking posterior assumptions that work well with both large and small datasets (Liddell & Kruschke, 2018).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science, 2*(1), 77–101. https://doi.org/10.1177/2515245918823199
Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories* (1st ed.). Cambridge University Press.
Gelman, A., & Robert, C. P. (2013). "Not only defended but also applied": The perceived absurdity of Bayesian inference. *The American Statistician, 67*(1), 1–5.
Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review, 48*(3), 398–407. https://doi.org/10.2307/2095231
Jones, A. L., Jaeger, B., & Schild, C. (2021). No credible evidence for links between 2D: 4D and COVID-19 outcomes: A probabilistic perspective on digit ratio, ACE variants, and national case fatalities. *Early Human Development, 152*, 105272. https://doi.org/10.1016/j.earlhumdev.2020.105272
Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2 ed.). Academic Press.
Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science, 1*(2), 270–280. https://doi.org/10.1177/2515245918771304
Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4
Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *Journal of Machine Learning Research, 18*(1), 430–474.
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963
Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009
McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician, 73*(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253
Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*(2), 103–115. https://doi.org/10.1086/288135

Petersen, M. B., Sznycer, D., Sell, A., Cosmides, L., & Tooby, J. (2013). The ancestral logic of politics: Upper-body strength regulates Men's assertion of self-interest over economic redistribution. *Psychological Science, 24*(7), 1098–1103. https://doi.org/10.1177/0956797612466415

Richardson, T. (2020). Height is associated with more self-serving beliefs about wealth redistribution. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2020.12.001

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science, 2*, Article e55. https://doi.org/10.7717/peerj-cs.55

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*(35), 15073–15078. https://doi.org/10.1073/pnas.0904312106

Smith, G. (2018). Step away from stepwise. *Journal of Big Data, 5*(1), 32. https://doi.org/10.1186/s40537-018-0143-6

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4