

The influence of the applicants' gender on the modeling of a peer review process by using latent Markov models

LUTZ BORNMANN,^a RÜDIGER MUTZ,^a HANS-DIETER DANIEL^{a,b}

^a Professorship for Social Psychology and Research on Higher Education, ETH Zurich,
Zähringerstr. 24, CH-8092 Zurich, Switzerland

^b Evaluation Office, University of Zurich, Zurich, Switzerland

In the grant peer review process we can distinguish various evaluation stages in which assessors judge applications on a rating scale. BORNMANN & AL. [2008] show that latent Markov models offer a fundamentally good opportunity to model statistically peer review processes. The main objective of this short communication is to test the influence of the applicants' gender on the modeling of a peer review process by using latent Markov models. We found differences in transition probabilities from one stage to the other for applications for a doctoral fellowship submitted by male and female applicants.

Introduction

In the grant peer review process we can distinguish various evaluation stages in which assessors judge applications on a rating scale. There are four main reasons, why a latent Markov model offers a fundamental good opportunity to model statistically peer review processes [BORNMANN & AL., 2008]: First, peer review processes base on categorical judgements (e.g., "award" or "no award"), which can be easily interpreted as frequencies in a statistical framework. In latent Markov models the input and the estimated parameters are probabilities, which strongly facilitate the interpretation of the results. Second, latent Markov models allow to model multi-stage peer review processes as a transformation of such probabilities ("award", "possible award" ...) over time, which gives some insight into how judgements in the process come about. Third, peer review processes are not free from random measurement errors: the different evaluation stages might slightly vary in scoring the material. With *latent Markov* models it is possible to generate ideal rating scale categories of an underlying quality dimension, so called latent classes, which, in contrast to the categories used by the assessors in assessing applications, (e.g., "award", "possible award", and "no award") give error-free measurement.

Received September 30, 2008

Address for correspondence:

LUTZ BORNMANN

E-mail: bornmann@gess.ethz.ch

0138-9130/US \$ 20.00

Copyright © 2009 Akadémiai Kiadó, Budapest

All rights reserved

BORNMANN & AL. [2008] analyzed 1954 applications for a doctoral or post-doctoral fellowship of the Boehringer Ingelheim Fonds (B.I.F.), assessed in three stages (first: evaluation by an external reviewer; second: internal evaluation by a staff member; third: final decision by the B.I.F. Board of Trustees) (see here BORNMANN & DANIEL, 2005A). The results of latent Markov models show that an application only has a chance of approval if it was recommended for support in the first evaluation stage. The main objective of this study is to test the influence of the applicants' gender on the modeling of a peer review process by using latent Markov models. Non-scientific properties of the applicants (such as gender) are functionally irrelevant for the progress of science, and to the extent that they are used as explicit or hidden criteria in the evaluation of scientific work, the principle of universalism is being abridged [COLE, 1992; MARSH & AL., 2008]. In the last few years a series of studies have been made analyzing the influence of non-scientific statuses of applicants on judgements in the peer review (see for example [LEDIN & AL., 2007]). Our findings of a meta-analysis of 21 studies (40 effect sizes for grants, 26 for fellowship applications) show that all in all, among applicants men have statistically significantly greater odds of receiving funding than women by about 7% [BORNMANN & AL., 2007]. In an extension and re-analysis of our data, the results reveal that gender differences in favor of men are larger for fellowship applications than for grant applications. Indeed, there are no statistically significant gender differences at all for grant applications [MARSH & AL., IN PRESS].

Methods and results

A detailed description of the latent Markov modeling approach is given by BORNMANN & AL. [2008]. We used the statistical program package PANMARK (PANel analysis using MARKov chains, [VAN DE POL & AL., 2000] for the estimation of the latent Markov models.

Comparison of two models

To test the influence of the applicants gender on the ratings in the B.I.F. peer review process, we calculated two latent Markov models separately for applications for a doctoral and post-doctoral fellowship: while the first model (M_1) does not differentiate between male and female applicants, the second model (M_2) allows for this difference. Table 1 shows for applications for a doctoral fellowship (p_{LR}^{boot}) that the LR-differences ($M_1 - M_2$) tested by χ^2 -statistics are statistically significant ($M_1 - M_2$: $\Delta LR = 76.65 - 50.96 = 25.69$ $p < 0.01$, $df = 39 - 31 = 8$). The result makes it apparent that doctoral applicants gender influences the outcome of the B.I.F. peer review process. For applications for a post-doctoral fellowship on the other hand (Table 1) the null hypothesis of no gender differences cannot be rejected. The LR does not drop

statistically significantly from 40.75 (M_1) to 36.38 (M_2) ($\Delta LR = 4.37$ $p > 0.01$, $df = 6$). Hence, the outcome of the process for this application group does not appear to be significantly influenced by gender (see here [BORNMANN & DANIEL, 2005B]).

Table 1. Likelihood-ratio test statistic for models M_1 (gender differences are *not* considered in the model) and M_2 (gender differences are considered in the model) ($n = 1954$, 75.4% applications for a doctoral fellowship, 24.6% applications for a post-doctoral fellowship)

| Model | df | LR | p_{LR}^{boot} |
|---|----|-------|-----------------|
| Applications for a doctoral fellowship | | | |
| M_1 | 39 | 76.65 | 0.00 |
| M_2 | 31 | 50.96 | 0.02 |
| Applications for a post-doctoral fellowship | | | |
| M_1 | 38 | 40.75 | 0.35 |
| M_2 | 32 | 36.38 | 0.26 |

Notes. df = degrees of freedom, LR = Log-likelihood ratio,
 p_{LR}^{boot} = bootstrapped calculated probability for log-likelihood ratio

Latent transition probabilities

As the comparative model has shown a significant influence of the gender on the B.I.F. peer review process for applications submitted by applicants for a doctoral, but *not* for a post-doctoral fellowship, Table 2 shows the latent transition probabilities (M_2) of ratings only for male and female applicants who submitted an application for a *doctoral* fellowship. For the transition ($t_1 - t_2$) from the first evaluation stage (external reviewer) to the second evaluation stage (staff member) applications of males have a greater chance ($\tau_{11} = 0.67$) than females to stay in the first latent class (“award”; $\tau_{11} = 0.59$). Accordingly, applications of females that are in the first latent class (“award”) at the first evaluation stage change more frequently to the second (“possible award”, $\tau_{12} = 0.30$) or third (“no award”, $\tau_{13} = 0.11$) latent class than applications of males (“possible award”, $\tau_{12} = 0.23$; “no award”, $\tau_{13} = 0.10$). Similar differences in transition probabilities for males and females can be seen for the transition from the second (staff member) to the third (Board of Trustees) evaluation stage of the B.I.F. peer review process ($t_2 - t_3$). Accordingly, males have a greater chance of remaining in the first latent class (“award”) – and so of being supported in the end – when progressing to the second (staff member) and third (Board of Trustees) evaluation stages than females. As with females, however, males have little chance of support if they have not already gained the “award” rating at the first evaluation stage (external reviewer).

However, the transition probabilities in Table 2 for the transition to the second evaluation stage (staff member) also indicate a countervailing gender influence in the assessment of B.I.F. applications. For applications of females ($\tau_{32} = 0.36$) there is a greater probability in this transition of changing from the third (“no award”) to the second (“possible award”) latent class than for applications of males ($\tau_{32} = 0.27$).

Correspondingly, applications of males ($\tau_{33} = 0.73$) that at the first evaluation stage (external reviewer) are in the third latent class (“no award”) are more frequently in the third latent class at the second evaluation stage (staff member) than applications of females ($\tau_{33} = 0.63$). An improvement in ratings from “no award” to “possible award” in the transition from the first to the second evaluation stage is more probable for female applicants than for male applicants.

Table 2. Estimated latent transition probabilities (and standard errors) for male and female applicants who submitted an application for a doctoral fellowship ($n = 1474$, row percent)

| Gender | Latent class (t_1) | Latent transition probabilities τ 's | | | | | | |
|--------|------------------------|---|----------------|----------------|------------------------|---------------------|----------------|----------------|
| | | From t_1 to t_2 | | | Latent class (t_2) | From t_2 to t_3 | | |
| | | Class 1 | Class 2 | Class 3 | | Class 1 | Class 2 | Class 3 |
| Male | 1 | 0.67 (0.03) | 0.23 (0.03) | 0.10 (0.02) | 1 | 0.23 (0.04) | 0.60 (0.06) | 0.17 (0.04) |
| | 2 | 0.00 (n.e.) | 0.61 (0.06) | 0.39 (0.06) | 2 | 0.00 (n.e.) | 0.21 (0.04) | 0.79 (0.04) |
| | 3 | 0.00 (n.e.) | 0.27 (0.04) | 0.73 (0.04) | 3 | 0.00 (n.e.) | 0.00 (n.e.) | 1.00 (n.e.) |
| Female | 1 | 0.59 (0.04) | 0.30 (0.04) | 0.11 (0.02) | 1 | 0.07 (0.04) | 0.67 (0.07) | 0.26 (0.05) |
| | 2 | 0.00 (n.e.) | 0.45 (0.06) | 0.55 (0.06) | 2 | 0.00 (n.e.) | 0.20 (0.04) | 0.80 (0.04) |
| | 3 | 0.01 (0.03) | 0.36 (0.05) | 0.63 (0.05) | 3 | 0.00 (n.e.) | 0.00 (n.e.) | 1.00 (n.e.) |

Notes. n.e. = standard error can not be calculated because of bounded parameters (0, 1). The response probabilities and class proportions are not shown, because they are quite similar to the model for the whole data see BORNMANN & AL., 2008.

Discussion

Using latent Markov models (see [BORNMANN & AL., 2008]) we examined the influence of applicants gender on the B.I.F. peer review process. The findings on the effect of gender on the B.I.F. peer review process are heterogeneous. The comparison of two latent Markov models (M_1 does not differentiate between male and female applicants; M_2 differentiates between both applicant groups) shows that applicants' gender has a statistically significant influence on the outcome of the B.I.F. peer review process in applications for a doctoral fellowship. No statistically significant influence could be found for applications for a post-doctoral fellowship. Therefore, we have only analyzed the latent transition probabilities for applications for a doctoral fellowship. The findings show that, although both males and females have little chance of funding if they have not already obtained the “award” rating from the external reviewer at the first evaluation stage, males however have a distinctly greater chance than females of remaining in the “award” category (and being funded) in the transitions from the first (external reviewer) to the second (staff member) and from the second to the third

(Board of Trustees) evaluation stage. The transition probabilities also show, however, a contrary gender effect on the assessment of B.I.F. applications: an improvement in the rating from “no award” – given by the external reviewer at the first evaluation stage – to “possible award” (staff member) is more likely for female applicants than male applicants.

NIEVA & GUTER [1980] also report similar contradictory gender effects in a research summary on the evaluation of the qualifications and performance of men and women in the work environment. “Bias ... appears to work in both directions. Competent males are rated more positively than equally competent females, while incompetent males are rated lower than equally incompetent females. This pattern of results can be reconciled by the notion of sex-role congruence. Because success at most demanding situations or occupations is generally expected of males and not of females, unsuccessful females are not as heavily penalized as unsuccessful males, from whom more is expected; however, females are not rewarded for success in the same way that males are” (p. 273).

*

We would like to thank Dr. Rolf Langeheine, retired professor of the IPN – Leibniz Institute for Science Education Kiel (Germany) for his helpful comments on estimating the latent Markov models.

References

- BORNMANN, L., DANIEL, H.-D. (2005A), Criteria used by a peer review committee for selection of research fellows. A Boolean probit analysis. *International Journal of Selection and Assessment*, 13 (4), 296-303.
- BORNMANN, L., DANIEL, H.-D. (2005B), Selection of research fellowship recipients by committee peer review. Analysis of reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, 63 (2) : 297–320.
- BORNMANN, L., MUTZ, R., DANIEL, H.-D. (2007), Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1 (3) : 226–238.
- BORNMANN, L., MUTZ, R., DANIEL, H.-D. (2008), Latent Markov modeling applied to grant peer review. *Journal of Informetrics*, 2 (3) : 217–228.
- COLE, S. (1992), *Making Science. Between Nature and Society*, Cambridge, MA, USA, Harvard University Press.
- LEDIN, A., BORNMANN, L., GANNON, F., WALLON, G. (2007), A persistent problem. Traditional gender roles hold back female scientists. *EMBO Reports*, 8 (11) : 982–987.
- MARSH, H. W., JAYASINGHE, U. W., BOND, N. W. (2008), Improving the peer-review process for grant applications – reliability, validity, bias, and generalizability. *American Psychologist*, 63 (3) : 160–168.
- MARSH, H. W., BORNMANN, L., MUTZ, R., DANIEL, H.-D., O'MARA, A. (in press), Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*.
- NIEVA, V. F., GUTER, B. A. (1980), Sex effects on evaluation. *Academy of Management Review*, 5 (2) : 267–276.
- VAN DE POL, F., LANGEHEINE, R., DE JONG, W. (2000), *PANMARK 3: User's Manual: PANel Analysis Using MARKov Chains; A Latent Class Analysis Program*, Voorburg, Netherlands Central Bureau of Statistics.