SNSF Report

Chiara Gilardi, Leslie O'Bray, Carla Schaerer and Tommaso Portaluri

13 June 2018

Contents

Introduction

The Swiss National Science Foundation (SNSF) is a research funding agency which disseminates yearly, on behalf of the Swiss Government, billions of CHF to the best researchers in Switzerland. This report contains a statistical analysis performed on three data sets provided by SNSF, containing information on the applications for funding received in 2016, the corresponding grades given by both internal and external evaluators, and whether the application was funded or not.

The SNSF evaluation procedure is a multi-step process (involving external reviewers, internal referees, and an internal board) which takes into consideration both the track record of the applicant and the quality of the project (see Appendix for a more detailed description of the evaluation procedure).

The analysis performed for SNSF had a two-fold aim, corresponding to the following two research questions:

(1) Is gender bias occurring at any stage of the SNSF evaluation process? Is the gender of the main applicant influencing the rating of the application?

(2) To what extent the different steps of the evaluation and the different criteria within each step determine the final funding decision?

Several studies (Witteman et al., 2017; Solans-Domenech et al., 2017) have shown that female applicants' projects get higher score when the application is blinded. Moreover, female applicants receive usually higher grades for projects and lower grades for track record. Hence, after investigating the gender dimension to identify possible biases in the evaluation procedure, the focus of the analysis will be the relative importance of of the criteria for funding (applicant's track record vs. quality of the proposal) and, also, of each step of the evaluation procedure (which opinion is more likely to determine the final decision - the external reviewer or the internal reviewer?). Possible interactions between the gender dimension and the second research question will also be investigated (for instance, by taking into account also the gender of evaluator or the percentage of female referees).

Data Description

We have three data sets: Applications, External Reviewers and Internal Referees. They contain respectively information about the SNSF project funding applications, the evaluation of the applications by external peer reviewers and the evaluation of the proposals by external the internal referee and co-referee (when available). For a full description of the data & variables, please see the Appendix.

Cleaning the Data

We made the following decisions while cleaning the data:

- We only worked with complete applications, i.e. projects for which we have information from all the three data sets.
- To avoid a temporal trend, we are only considering application from 2016.
- In both the external and internal step, we encountered applications which had several reviews per application. For the sake of our analysis, in these scenarios we computed the mean grade for each criteria, so that each application had a "single" score for each criteria assessed on. In all instances of combining scores, we did not round until the very end, and rounded up if the value was ≥ 0.5, otherwise we rounded down.
- Since an application had different numbers of reviewers, we introduced a new variable, PercentFemale, which calculated the percent of female reviewers out of all reviewers of a single application (ranging from 0 to 1).

- All applications with a grade were converted to an ordinal factor, with levels from 1 (lowest) to 6 (highest).

We also made some additional decisions given the specific datasets:

Applications

- We only consider the MainDiscipline2 (a factor with 21 levels), rather than MainDiscipline (a factor 118 levels) to improve interpretability.
- There is one application for which we do not know the gender of the applicant, and therefore we decided to omit that observation from the analysis.
- We will also not consider the variable "CallTitle" because we do not think it has value to the model. We also will not consider "Professorship" or "AcademicAge" due to the fact that there are a considerable number of NA's on those variables (around 93% of the observations).

Type	Frequency
Assistant professor with tenure track	102
Assistant professor without tenure track	54
Associate professor	237
Full professor	512
Honorary professor or Titular professor	77
None	430
Professor at UAS / UTE	74
Visiting professor	4
NA	20037

External Reviewers

- We did not consider reviewer observations that did not give a grade or gave a grade of "0". Reviewers always have the option to choose not to consider or to give the grade "0" when reviewing an application. Some might be mistakes, in others cases there might be a conflict of interest, or they might be very ambivalent about the project.
- One of the questions evaluated in the applications is "Broader impact (forms part of the assessment of scientific relevance, originality and topicality)". For the time frame we are considering, in all the applications this grade was NA. Hence, we omit this variable from our model.
- ProposalCombined: We created a new variable to summarize the assessment of the scientific proposal in the external review step. This is a simple mean of the grade given for Suitability and Scientific Relevance. This helped to isolate the effect of the grade given to the scientific proposal, versus the applicant track record, as well as to ensure easier comparison with the internal review step.
- PercentFemale: As previously mentioned, we introduced a new variable calculating the percent of reviewers of each application that is female.

Internal Referees

- There were 22 observations (1 for the time frame we are dealing with) for which only demographic information was available, no grades were given. We decided to omit those observations.
- We decided not to consider the Referee role as a variable in our model, as the majority of the evaluations has only one referee.

Type	Frequency
Applicant	2
Explicit inclusion	20
Recusal	8
Referee	15766
Second referee	870
NA	1276

- PercentFemale: As previously mentioned, we introduced a new variable calculating the percent of reviewers of each application that is female.

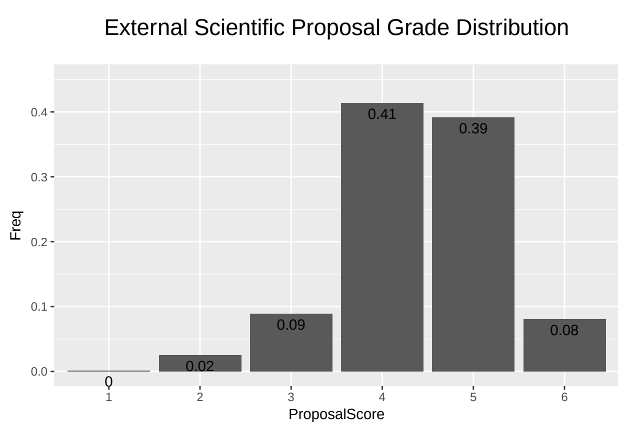
Exploratory Analysis

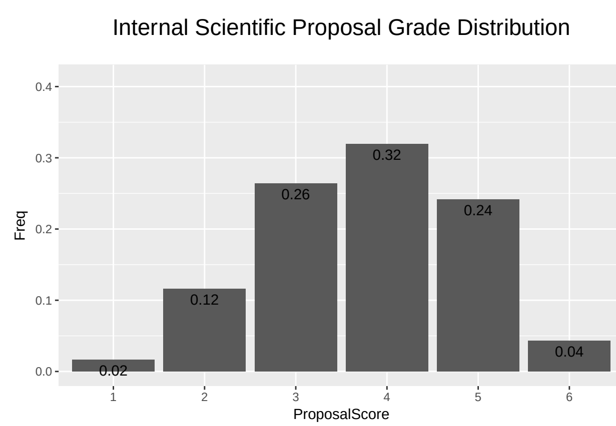
In our exploratory analysis, we discovered a few interesting insights that relate to the findings we will discuss from our formal analysis.

Comparison of the Distribution of Grades between the External & Internal Review Step

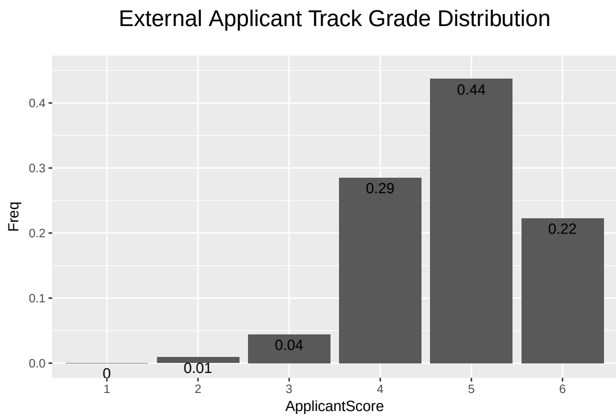
Since the external & internal step both assess candidates on the same criteria (the strength of the scientific proposal, and the strength of the applicant), on the same ordinal scale (from poor to outstanding), we were interested to see if the distribution of grades are the same. We would expect different distributions for the Overall Grade vs the Ranking, since those are measures are on different scales, but we may expect to see a similar distribution of grades for criteria that use the same measurement system. After combining the Suitability & Scientific Relevance grades given to a candidate in the external review step into a single grade, we can compare the average grade given for the Scientific Proposal in the two steps, as well as the grade given for the Applicant Track Record across the two steps.

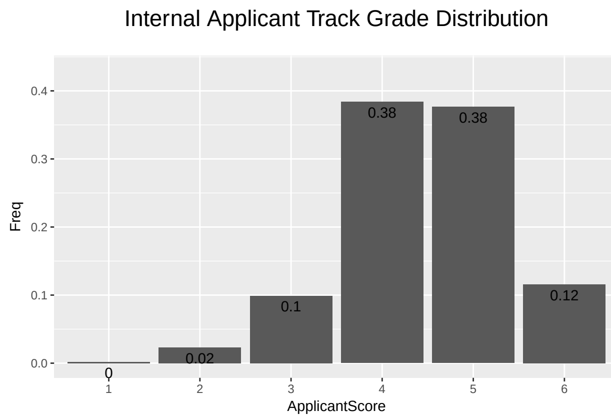
We see that the External Reviewers are more generous with their grades; for the strength of the Scientific Proposal, 48% of proposals are considered "excellent" or "outstanding", versus only 28% in the internal review step.





Similarly we see the same pattern with Applicant Track Record: 66% of Applicant Track records are considered "excellent" or "outstanding" by the External Reviewers, versus merely 50% by the Internal Reviewers.





After noticing this discrepancy, we wanted to quantify it. To assess the agreement between the reviewers assessing the same thing, we used Cohen's Kappa. Cohen's Kappa measures the proportion of agreement between two raters assessing something on an ordinal scale, accounting for the fact that there will always be some agreement by random chance. An important specification of Cohen's Kappa is the weight given to the measurements. If the external & internal reviewers both assessed the Applicant Track Record as "excellent", that would be considered full agreement. However, we want to allocate partial credit if the rating is a level close to it. We used a linear weight up to distance 2, and after that gave no credit. (In this example, if the first rater gave "excellent", and the second rater gave "very good", that would be considered a distance of one and be weighted by 0.8. If the second rater assessed the Applicant Track to be "good", which is a distance of two away from "excellent", that would be weighted as 0.6). Anything with a distance of 3 or more (in this example, if the second rater gave a rating of "average"), we allocated no weight, as the difference between average and excellent is quite large.

Cohen's Kappa takes a value from 0 to 1, with 0 signifying there is no more agreement than what one would expect from random chance, and 1 indicating perfect agreement. From this, we found that there was just "moderate" (kappa between 0.4-0.6) agreement for the grades given for the Applicant Track record between the internal and external reviewers when using the weighted kappa:

```
Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)

Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
                 lower estimate upper
unweighted kappa  0.21     0.24  0.28
weighted kappa    0.35     0.42  0.49

 Number of subjects = 1623
```

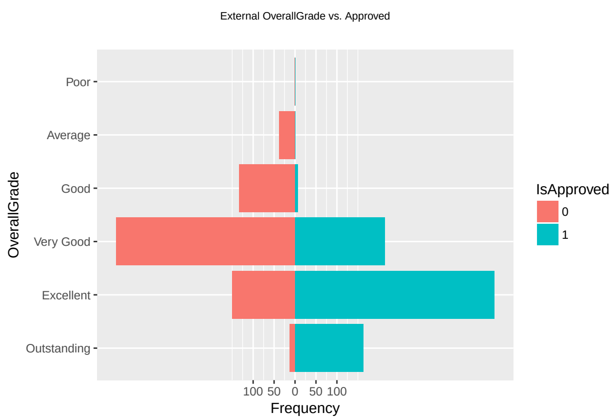
We observe the same moderate agreement when we look at the comparison of grades given to the scientific proposal (ProposalCombined and ProjectAssessment):

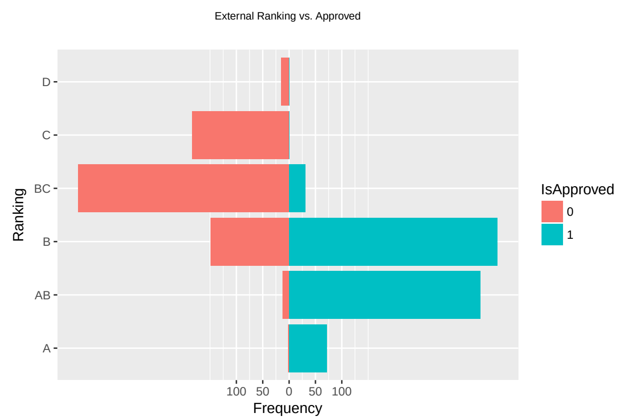
```
Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)

Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
                 lower estimate upper
unweighted kappa  0.17     0.19  0.22
weighted kappa    0.34     0.40  0.47

 Number of subjects = 1623
```

Impact of External & Internal Grade on Funding

Having observed a discrepancy is how grades are allocated between the external and internal step, we wanted to understand if this difference had a meaningful impact on the overall funding decision. We visualized the summary grade given to an application in the external and then internal step, and plotted how many applications within that grade was funded (color coral) and not funded (color blue). As we can see here, there are several applications in the external step that receive an OverallGrade of "excellent" or "outstanding" that ultimately are not approved. This highlights that not only do the Internal Reviewers give tougher grades in general than the external step, but they also consider some "excellent" and "outstanding" applications by the external reviewers to be not of the quality that deserves funding. This trend is true in all divisions and both genders, please refer to the appendix to see the specific graphic. On the other hand, applications with top grades in the Ranking almost always are approved.

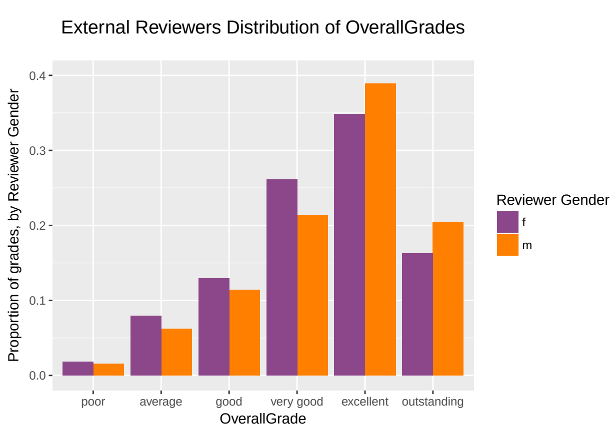
Our conclusion for this is that the internal step is very consequential, and the difference in the rating they give translates into differences in whether an application gets funded or not.

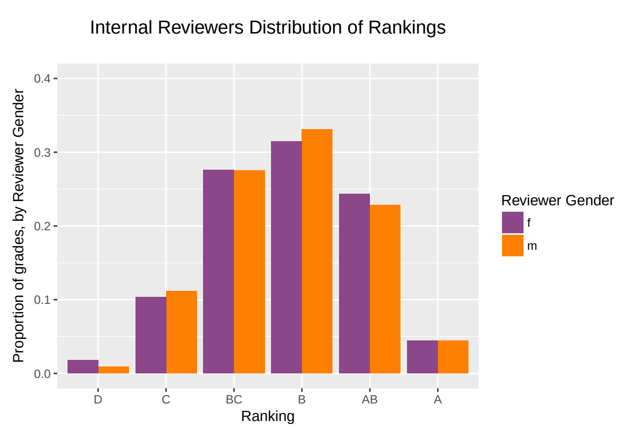




Distribution of Grades by the Gender of the Reviewer

The third interesting insight we found was when we investigated the impact of the gender of the person reviewing the application. We look at the relative frequencies of grades given by male and female reviewers, to applicants, regardless of gender. Within the external step in particular, we found that female reviewers give proportionally fewer "excellent" and "outstanding" grades compared to their male counterparts. Within the internal step, we did not notice a particular difference, though we will consider the impact of the gender of the reviewer more rigorously in our analysis.





Gender Bias

To see if gender has an influence in any of the steps of the evaluation process, we did several things. We wanted to check at an overall level, is there evidence of a gender bias in the funding process? And then within each the external and internal steps, does gender influence the grade applicants receive? For the former we conducted a χ^2 test of the independence. For the latter question, we took took the following two approaches:

(1) **Logistic Regression:** We first fit a logistic regression with the function glm in R, where we used IsApproved (a binary variable) as a response and demographic information of the applicant, project information and the given grades as predictors. The aim of this regression is to see if gender has an influence on the final decision from the perspective of each step.

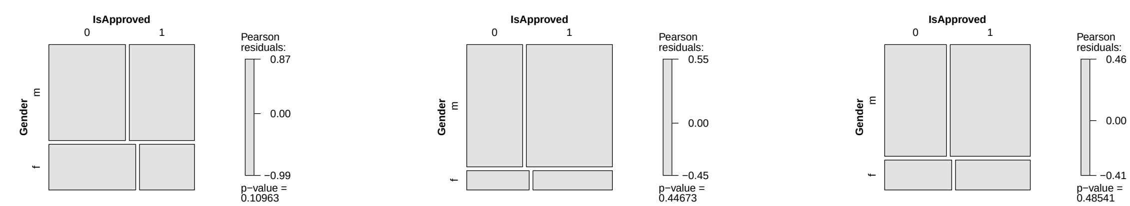
(2) **Ordinal Regression:** Next, as the final decision is determined by the different grades in the process, in order to see if gender has an influence on any of them, we fitted an Ordinal regression with the function clm of the package "ordinal" on each grade with demographic data and project information as predictors.

In the following sections we will review the analysis of gender bias at an overall level, as well as the analyses for the external, and then internal, steps of the process.

Gender Bias in the Funding Decision

To assess if there is a gender bias in the overall funding decision, we did a χ^2 test of independence between Gender and IsApproved, to see if there is a significant relationship these two categorical variables. A χ^2 test of independence at an organizational level would conclude these two variables are not independent (which itself only suggests an association, and not a causal relationship, between Gender and IsApproved), however, we decided the more appropriate test is to check for a relationship between Gender and IsApproved within each Division, since the funding decision is made within each Division, not at the organizational level.

From this analysis, we see that while all three Divisions have a lower acceptance rate for women than men, this difference is not significant at a 5% significance level. We therefore fail to find evidence of a gender bias in the overall funding decision.



External Review Step

We now proceed to investigate if there is a gender bias within the External Review Step.

Logistic Regression

The goal of this analysis is to see if gender has an influence on the final funding decision, based on the information provided in the external step.

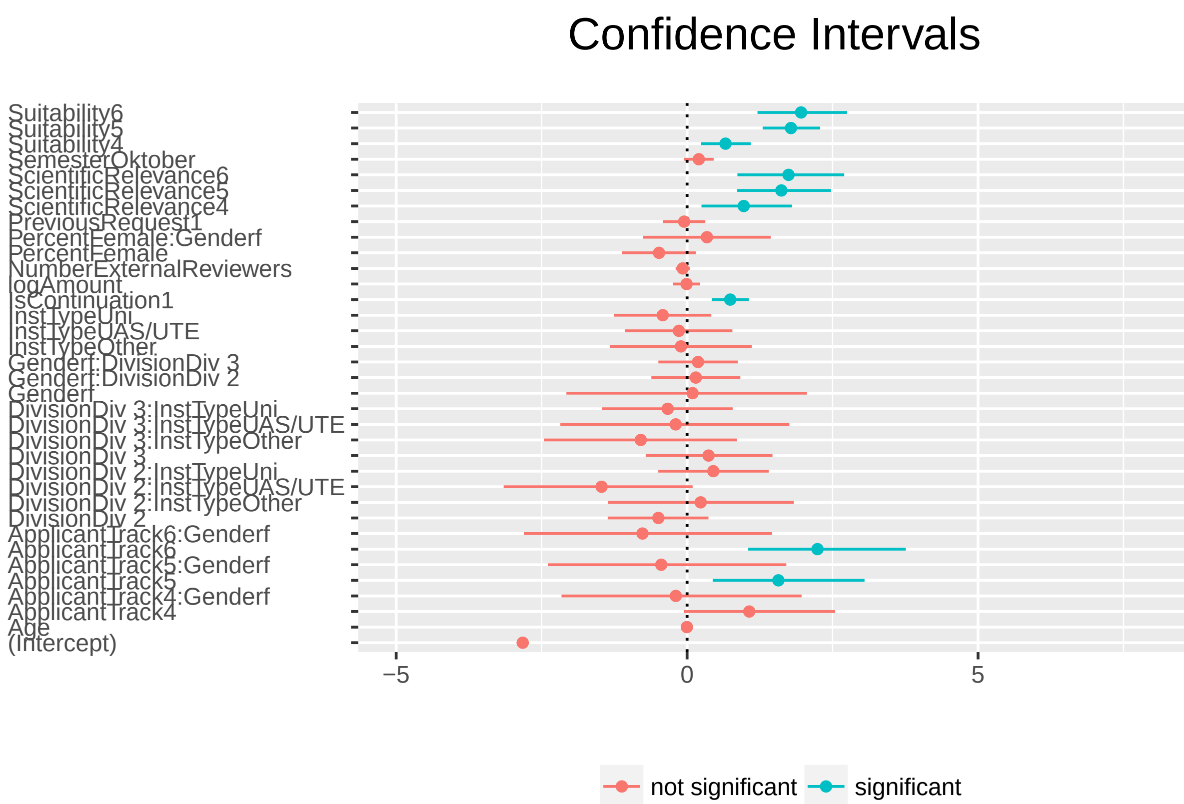
To perform the analysis, we combined in one data frame information about the applications (IsApproved, Age, Gender, Division, IsContinuation, PreviousRequest, InstType, log(AmountRequested), Semester) and about the grades given by the external reviewers (ApplicantTrack, ScientificRelevance, Suitability, OverallGrade, ProposalCombined, PercentFemale). We only had two continuous variables, and at this stage did not standardize them.

As almost no application with grade below “good” was approved, we decided to aggregate grades “poor”, “average” and “good” to avoid perfect separation problems. All grades are considered as factors.

We first fitted a logistic regression model with all the variables and the interactions between Gender and Division, PercentFemale and ApplicantTrack with the glm function in package *stats*. We also considered the interaction between InstType and Division. We didn’t considered OverallGrade as it is highly correlated with the grades of the Applicant Track and the scientific proposal. This model had a pseudo-R^2 value of 0.4225, indicating that this percent of the variation in the approval of the applications can be explained by the model. [1]

When selecting the variables with the AIC criteria in order to work with a small and effective model, we end up with the following predictors: ApplicantTrack, ScientificRelevance, Suitability, PercentFemale, Age, Gender, Division, IsContinuation, InstType, Semester, Division:InstType. No interactions with Gender were significant. The pseudo-R^2 for this model is 0.4203, i.e. this smaller model explains roughly the same variance of the data than the former one. None of them reveal that the information from the external reviewers explain the final decision correctly. This fact will be explored in more detail later on. [2]

In order to assess if gender is a relevant predictor for this regression, we computed the 95% confidence intervals for the coefficients of our model. We can see them in the following plot: those that are red include 0 and therefore they are not significant. Those in blue don’t include zero and so they are significant at a 5% significance level.



[1] We used the following computation of pseudo-R^2: $R^2 = \frac{1-(exp((D_{res}-D_{null})/n)}{1-exp(-D_{null}/n)}$

[2] See Appendix pg. for the summary result of this regression.

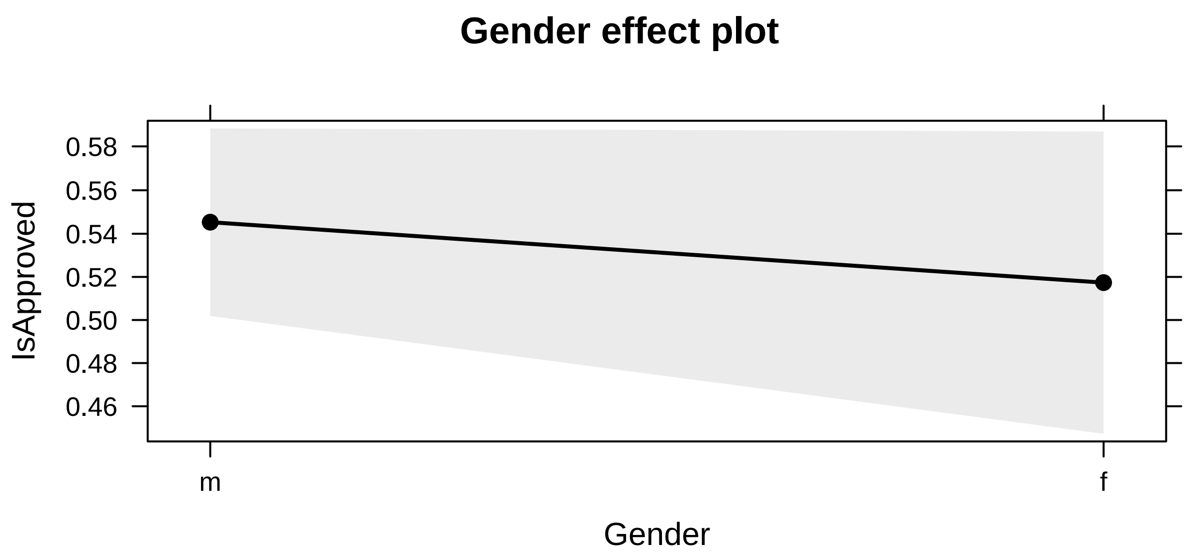
The confidence intervals which do not include zero are those for the variables: IsContinuation, Suitability (all levels from 4 to 6), ScientificRelevance (all levels from 4 to 6), and ApplicantTrack(for levels 5 and 6).

	OR	2.5 %	97.5 %
ApplicantTrack5	4.80	1.60	12.32
ApplicantTrack6	9.40	2.82	24.00
ScientificRelevance4	2.65	1.30	6.13
ScientificRelevance5	5.05	2.40	11.98
ScientificRelevance6	5.72	2.46	15.35
Suitability4	1.94	1.27	2.97
Suitability5	5.96	3.66	9.77
Suitability6	7.10	3.46	16.04
IsContinuation1	2.10	0.25	2.81

The interpretation of the exponentiated coefficients is as following: the odds is the $\frac{P(Approved=1|Grade=x)}{P(Approved=0|Grade=x)}$.

- The odds with respect to being approved or not given an Applicant Track grade of 5 is 4.8 times larger than the odds for the reference level
- The odds with respect to being approved or not given an Applicant Track grade of 6 is 9.4 times larger than the odds for the reference level
- The odds with respect to being approved or not given a Scientific Relevance grade of 4 is 2.65 times larger than the odds for the reference level
- The odds with respect to being approved or not given a Scientific Relevance grade of 5 is 5.05 times larger than the odds for the reference level
- The odds with respect to being approved or not given a Scientific Relevance grade of 6 is 5.72 times larger than the odds for the reference level
- The odds with respect to being approved or not given a Suitability grade of 4 is 1.94 times larger than the odds for the reference level
- The odds with respect to being approved or not given a Suitability grade of 5 is 5.96 times larger than the odds for the reference level
- The odds with respect to being approved or not given a Suitability grade of 6 is 7.1 times larger than the odds for the reference level
- The odds with respect to being approved or not given that the project is a continuation is 2.1 times larger than the odds for the reference level

Notice that the coefficient for gender is not significant, since its confidence interval (-2.074,2.059) contains zero. The point estimate for gender is 0.094 and this means that the odds with respect to being approved or not given that the applicant is a women is 2.99 times larger than the odds for a male applicant.



As we are interested in the effect of gender in each step of the evaluation process, we looked at the difference of the predicted probability of being accepted between male and female given the final model. Female have a probability 0.517, while male have 0.545. Although there is a small difference, the confidence intervals overlap and thus the difference is not significant.

Ordinal Regression

To look into the variables that influence the different grades in this part of the evaluation process, and see if the gender of the main applicant has an influence on it, we ran an ordinal regression with the function clm of the package ordinal in R, which fits cumulative link models (CLMs), for further details of this models see Appendix pg.

We did this for all grades given by the external reviewers:

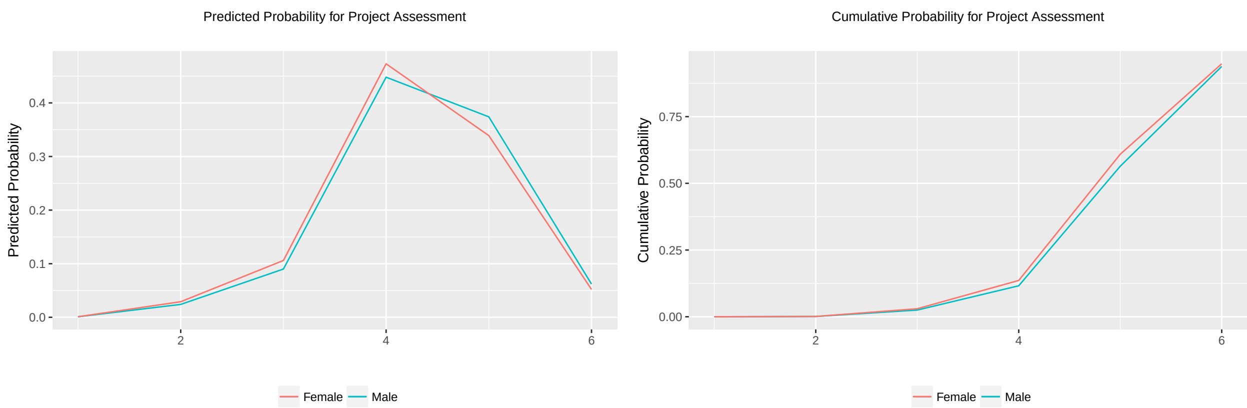
1. The average of the grades given to the scientific proposal (ProposalCombined)
2. The applicant grade (ApplicantTrack)
3. The Overall Grade.

(1) **Scientific Proposal (ProposalCombined)**: We fit a full model using the clm function with ProposalCombined as a response variable. We used all the predictors from the application and external reviewers datasets, along with varioius interactions, and then selected variables with the AIC criteria with the help of the drop1() function in R. From the output of the drop1() function, we manually excluded the variable whose omission would reduce the AIC the most, one by one. We ended up with a model with the following predictors: Gender, Division, PercentFemale, IsContinuation, InstType and log(AmountRequested). We fitted this same model without Gender and compared it using the likelihood ratio test from the anova() function to the same model with gender. We had a p.value of 0.1064133, meaning that for the grades given to the project, the model with gender is not significantly better than the model without gender. We thus conclude that we don't have evidence of gender having a substantial impact on the grade given to ProposalCombined.

Table 1: Predicted probabilities of getting different grades

	Male	Female	Difference
poor	0.001	0.001	0.000
average	0.024	0.029	-0.005
good	0.090	0.106	-0.015
very good	0.448	0.473	-0.024
excellent	0.374	0.339	0.035
outstanding	0.062	0.052	0.010

Overall the average difference is really small: 0.01483. This seems to suggest that there is no evidence of gender influencing the probability of achieving a certain grade. We also represented in the plots below the probability and cumulative probability curves of getting each grade for male and female: they follow more or less the same trend and the only difference, as we've seen from the table above, is that women are slightly more likely to get a "very good" rather than an "excellent" or "outstanding."



From the plot above we can see that gender is not significant, since its confidence interval includes zero. The other significant predictors are Division, the percentage of female reviewers, if the project is a continuation, the Institution type and the log(AmountRequested).

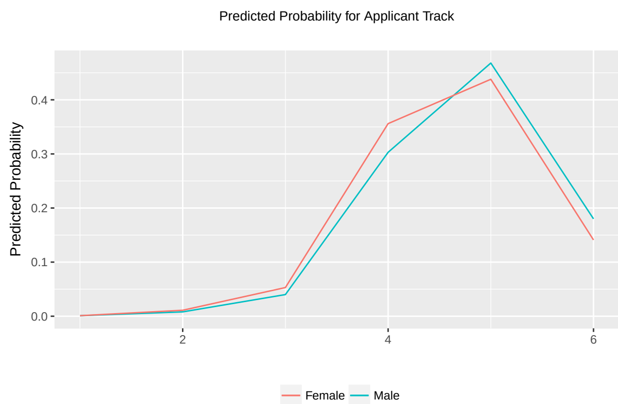
(2) **Applicant Track assessment**: Next, we fit another Ordinal regression with ApplicantTrack as a response variable. We used the following predictors: Gender, Division, PercentFemale, IsContinuation, InstType, log(AmountRequested) and the interaction between Gender and PercentFemale. Again we fitted the same model without Gender and compare it with the anova() function to the one with gender, we get a p.value of 0.0014313, meaning that for the grades given to the main applicant, gender needs to be considered in the model. In the next table we present part of the summary for this model, to see the full summary refer to the Appendix.

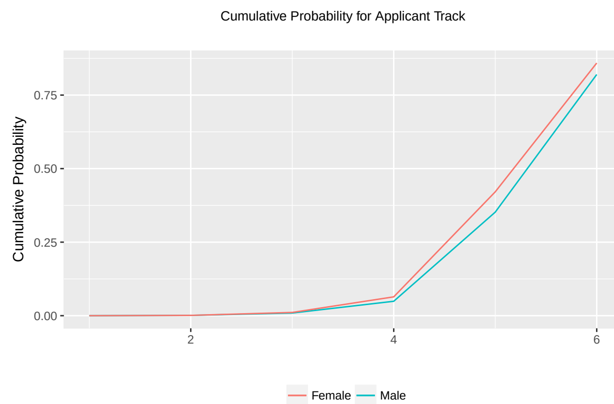
We computed the difference in predicted probability of getting a specific grade for both male and female.

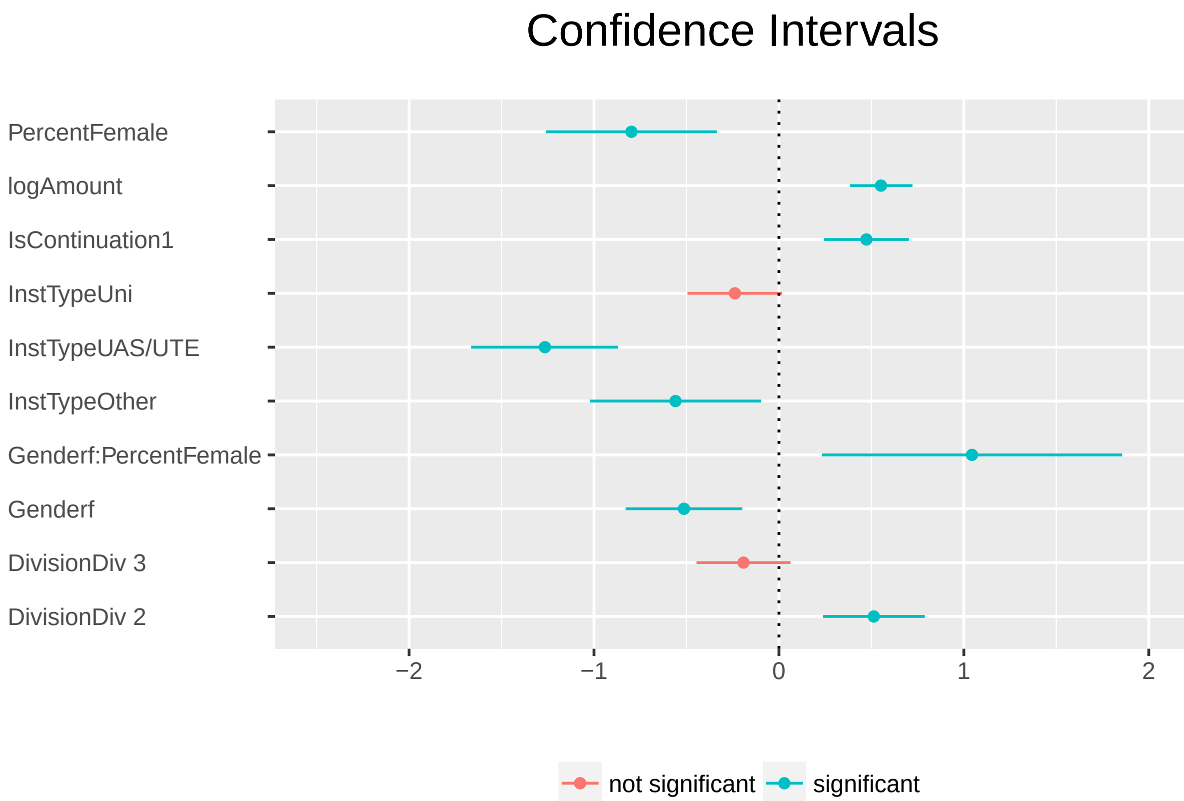
Table 2: Predicted probabilities of getting different Applicant Track grades

	Male	Female	Difference
poor	0.001	0.001	0.000
average	0.008	0.011	-0.003
good	0.040	0.053	-0.013
very good	0.303	0.356	-0.053
excellent	0.468	0.438	0.030
outstanding	0.180	0.141	0.039

In the table above, we see the predicted probability of getting each grade for both male and female and the difference between the two. The average difference of the cumulative probability is here 0.023, very close to zero. From the plot below we see that there is almost no difference between women and men probabilities. Although the difference between male and female is significant, it is a very large difference between male and female.







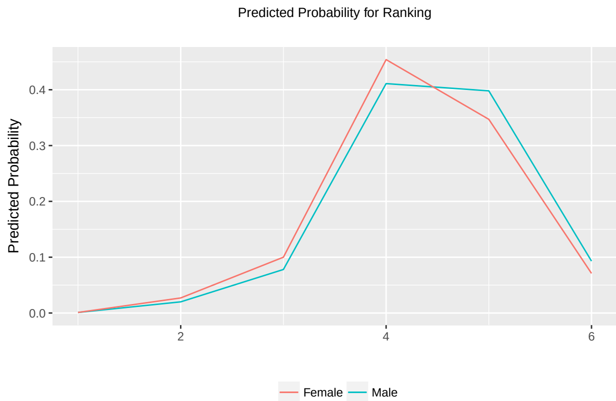
Notice that the confidence interval referring to gender does not include zero and so the corresponding coefficient is significant. The only variables that appear to be not significant in the determination of the Applicant Track grade, are the institution type and the Division. In this model, Division 1 is used as a base line, and from the results above, we can see that the only impact in grades, from the division point of view, is if the applicant comes from Division 2. Likewise, from the institution type point of view, there are different chances of reaching higher grades for applicants in institutions other than ETH and Uni.

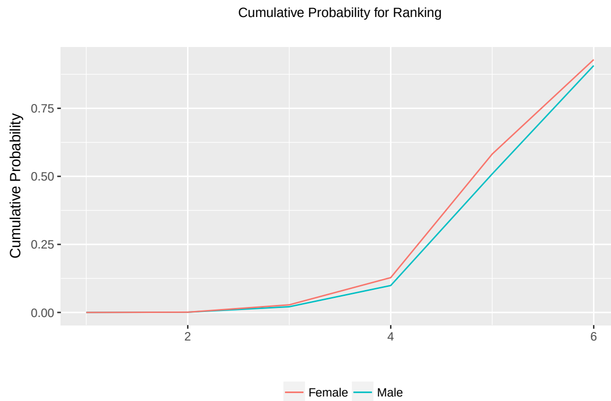
(3) **Overall Grade**: This last model has OverallGrade as a response and Gender, Division, PercentFemale, IsContinuation, InstType, PreviousRequest and log(AmountRequested) as predictors. We are not considering here the grades given to the applicant track record and to the project, as we just want to see the influence of the demographic data and the project information in each grade. A comparison of this model to the same one without gender suggests that gender is not significant: it's p.value is 0.11397.

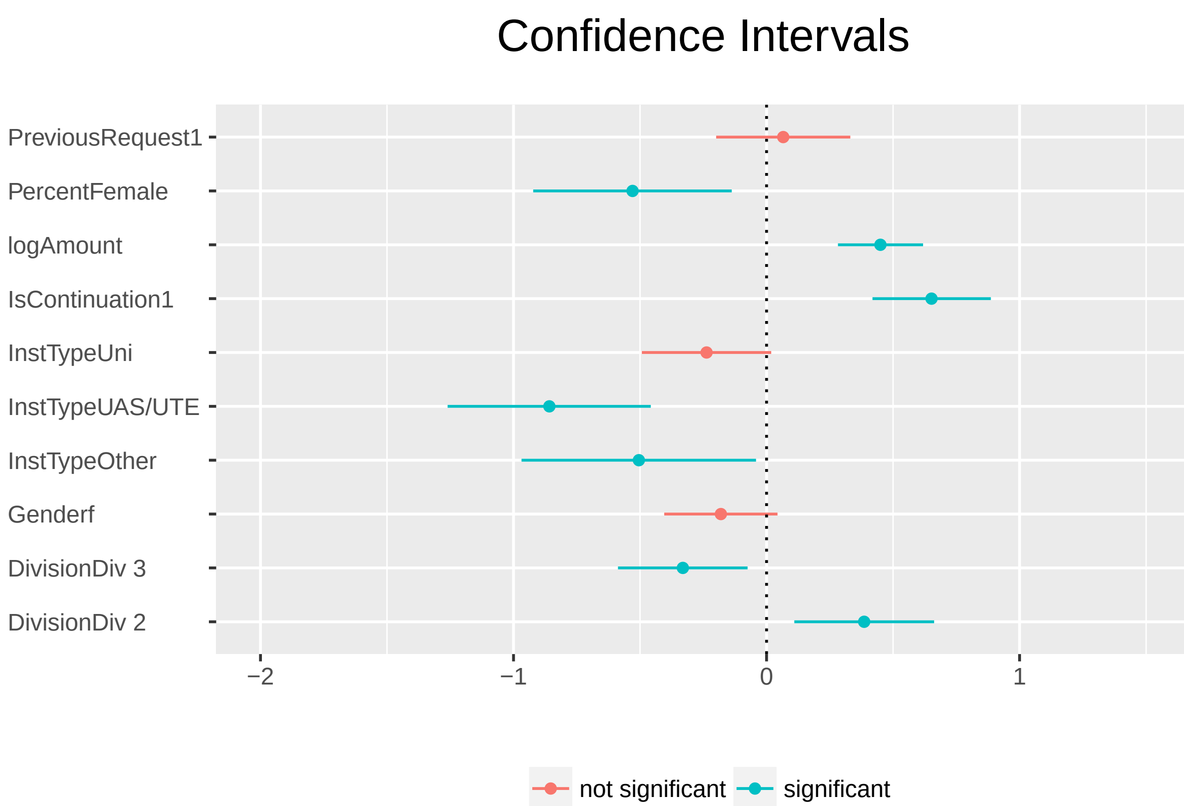
Table 3: Predicted probabilities of getting different Overall Grades

	Male	Female	Difference
poor	0.001	0.001	0.000
average	0.020	0.027	-0.007
good	0.078	0.100	-0.022
very good	0.411	0.454	-0.043
excellent	0.398	0.347	0.051
outstanding	0.093	0.071	0.022

The predicted probabilities of achieving certain grade for male and female is shown in the next table. The average difference of the cumulative probability is here as well close to zero (0.02417). Notice that the only difference is that female applicants are mole likely to get a “good” or “very good” grade rather than a “excellent” or “outstanding”, relative to male applicants.







Here gender seems to be not significant, since its confidence interval includes zero. However the difference between the upper bound and zero is really small. The other significant variables in this model are PercentFemale, log(AmountRequested), IsContinuation, the InstType and Division.

Here again gender seems to be not significant, since its confidence interval includes zero. However the difference between the upper bound and zero is really small. The other significant variables in this model are the percentage of female referees, PreviousRequest, and the institution type.

Internal Review Step

Logistic Regression

We now repeat the logistic regression for the internal step, to understand if gender has an influence on the final funding decision, based on the information provided in the internal step.

As explained before, in order to perform the analysis, we combined in a single data frame all the information about applications (IsApproved, Age, Gender, Division, IsContinuation, PreviousRequest, InstType, log(AmountRequested), Semester) and grades given by the internal referees (ApplicantTrack, ProjectAssessment, Ranking, PercentFemale).

We had again a perfect separation problem, due to the fact that there are very few approved applications with grades worse than "good". We aggregated the grades which were in category "poor", "average" and "good" in a unique category for both the applicant track record and the project assessment. All grades are considered as factors. Since the proportion of female referees who evaluate the application takes only values 0, 0.5 and 1, we consider it as a factor with 3 levels.

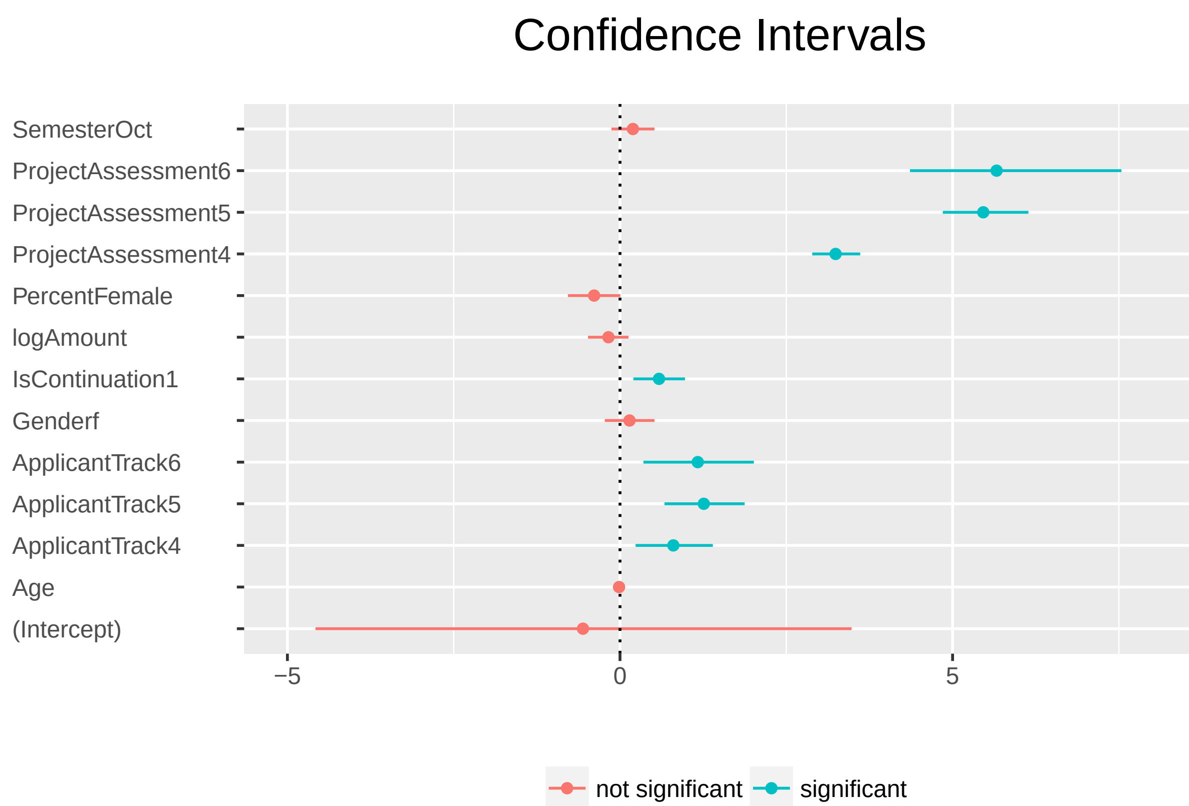
We didn't standardized the continuous variables, in order to be able to interpret the coefficients estimated from the model.

We first used the glm function in R to fit a full model with all the available variables and the interactions between Gender and Division, Gender and PercentFemale, and Gender and ApplicantTrack. We also considered the interaction between InstType and Division. We didn't include the Ranking grades into the model, since they are highly correlated with the single grades for the applicant and the project. We achieved a pseudo-R^2 value of 0.6914, meaning that the variation in the binary variable Y (approved or not) can be explained for more than half by this model including only the internal step information.

Then, we did variable selection using the AIC criteria again with the drop1() function, in order to obtain a small and effective model as we did for the external step. The remaining predictors are Gender, Age, Semester, IsContinuation, PercentFemale, ApplicantTrack, ProjectAssessment and log(AmountRequested). Interactions were removed from the model because they were not significant.

The pseudo-R^2 for this model is 0.686 , i.e. this reduced model explains basically the same proportion of variance of the data as the previous model. Even if we removed some variables, the model still explains almost 70% of the variation of the variable IsApproved. It therefore seems that the internal grades are significant predictors for the final funding decision.

In order to assess if gender is a relevant predictor for this regression, we computed the 95% confidence intervals for the coefficients of our model. We can see them in the following plot: those that are red include 0 and therefore they are not significant. Those in blue don't include zero and so they are significant at a 5% significance level.



The confidence intervals which don't include zero are those for the variables: IsContinuation, ApplicantTrack (all levels from 4 to 6) and ProjectAssessment (all levels from 4 to 6). Notice that the coefficient for gender is not significant, since its confidence interval (-0.2284, 0.5191) contains zero. The point estimate for gender is 0.1433 and this means that the odds with respect to being approved or not given that the applicant is a woman is 1.1541 times larger than the odds for a man (which is the reference level).

In order to interpret the coefficient estimates, we exponentiated them and we obtained the following values:

Table 4: Odd Ratio and corresponding confidence intervals, Internal Logistic Regression

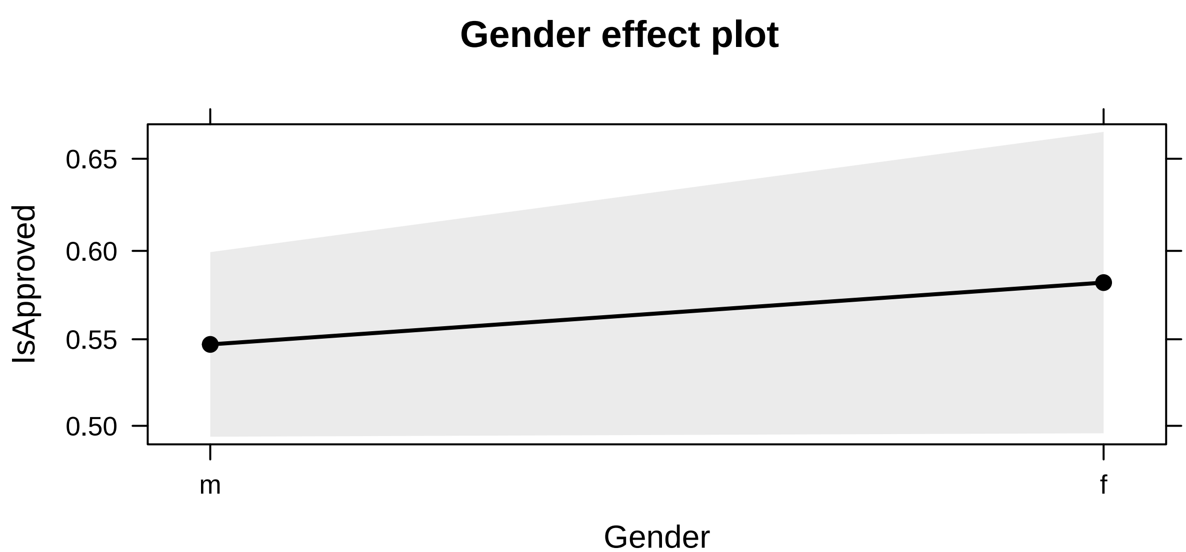
	OR	2.5 %	97.5 %
IsContinuation1	1.80	1.30	6.13
ApplicantTrack4	2.23	2.46	15.35
ApplicantTrack5	3.52	1.27	2.97
ApplicantTrack6	3.22	3.66	9.77
ProjectAssessment4	25.57	3.46	16.04
ProjectAssessment5	235.39	0.39	1.10
ProjectAssessment6	287.64	0.98	1.01

- The odds with respect to being approved or not given that the project is a continuation is 1.8 times larger than for a project which is not a continuation, keeping all other variables constant;
- The odds with respect to being approved or not given that the Applicant Track grade is 4 is 2.23 larger than the odds for the reference level (Applicant Track grade from 1 to 3), keeping all other variables constant;
- The odds with respect to being approved or not given that the Applicant Track grade is 5 is 3.52 larger

than the odds for the reference level (Applicant Track grade from 1 to 3), keeping all other variables constant;

- The odds with respect to being approved or not given that the Applicant Track grade is 6 is 3.22 larger than the odds for the reference level (Applicant Track grade from 1 to 3), keeping all other variables constant;
- The odds with respect to being approved or not given that the Project Assessment grade is 4 is 25.57 larger than the odds for the reference level (Project Assessment grade from 1 to 3), keeping all other variables constant;
- The odds with respect to being approved or not given that the Project Assessment grade is 5 is 235.39 larger than the odds for the reference level (Project Assessment grade from 1 to 3), keeping all other variables constant;
- The odds with respect to being approved or not given that the Project Assessment grade is 6 is 287.64 larger than the odds for the reference level (Project Assessment grade from 1 to 3), keeping all other variables constant.

As said before, our focus is the effect of gender in each step of the evaluation process. We checked the difference in the predicted probability of being accepted between male and female: women have probability 0.582 of the project being approved, while men have 0.547. Surprisingly, the predicted probability of being funded seems to be higher for female applicants, even if the difference is too small to be relevant. We can clearly see that from the Gender effect plot, where the confidence intervals are overlapping and the line is almost horizontal. From this initial analysis we can say that it doesn't seem that referees are biased against women.



Ordinal Regression

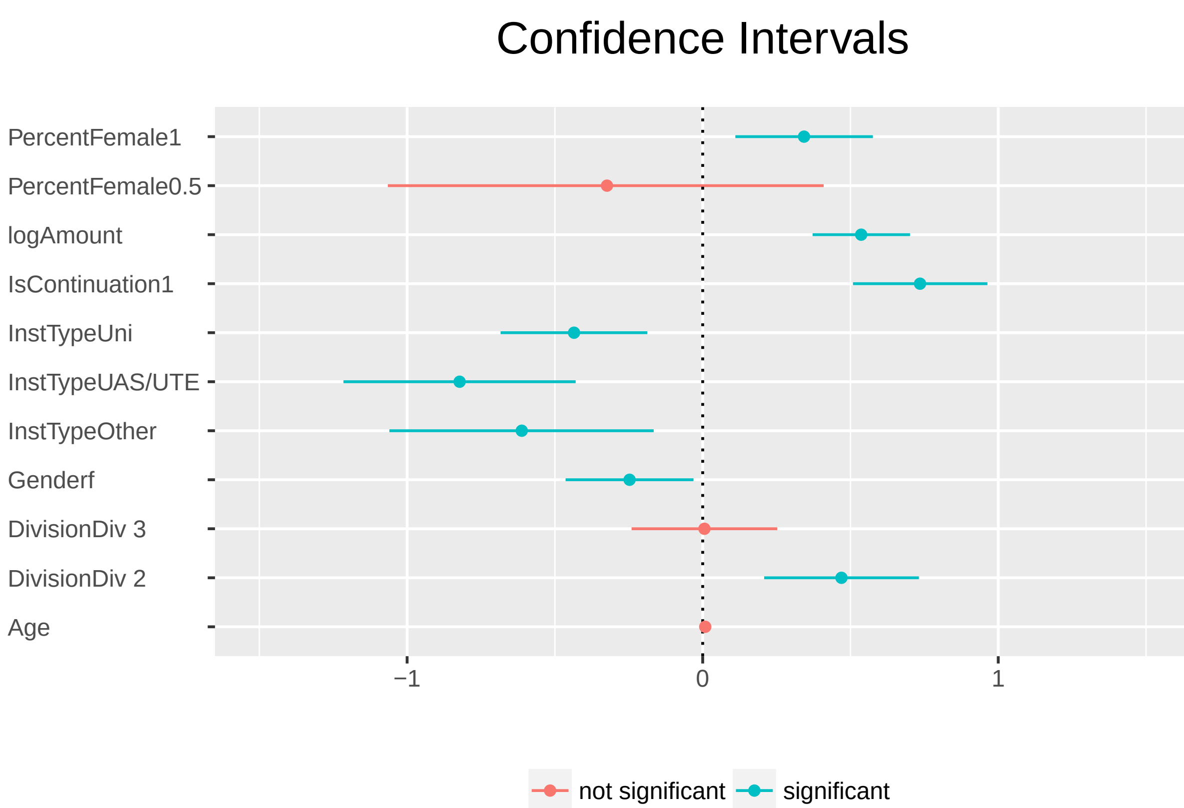
To see which variables influence the different grades in the second step of the evaluation process and check if the gender of the main applicant has an influence on it, we ran an ordinal regression with the function clm of the package *ordinal* in R.

We did this for all grades given by the internal reviewers:

1. The grades given to the scientific proposal (ProjectAssessment)
2. The applicant grade (ApplicantTrack)
3. The Ranking Grade.

(1) **Scientific Proposal (ProjectAssessment)**: We fitted the full model with ProjectAssessment as response variable, using the clm function. Then we selected the significant variables with the AIC criteria with the help of the drop1() function in R. We performed a stepwise procedure by hand, removing each time the variable that minimized the AIC most. We ended up with a model with the following predictors: Gender, Division, PercentFemale, Age, IsContinuation, InstType and log(AmountRequested). We also fitted the same model without Gender (using the clm function again) and compared it to the previous one using the likelihood ratio test implemented in the anova function: we got a p-value of 0.0249831, meaning that Gender seems to be a significant predictor for the grades given to the project by the internal referees.

Below are the predicted confidence intervals for all the coefficients in our model: all the variables are significant (except for one level of Division and one for PercentFemale). Notice that the upper bound of the gender confidence interval is really close to zero. From this results, we cannot say that there is clear evidence of gender bias.



Since we have gender as a predictor in our model, we can show the effect it has on the grades. For this purpose, we estimated the predicted probability of falling in the different categories for each gender and the difference between men and women. The result is presented in the following table.

Overall the average difference is really small: 0.02067. Even though gender is a significant predictor, we can see from the predicted probabilities that gender does not have a meaningful impact on the probability of achieving a certain grade.

Table 5: Predicted probabilities of getting different Scientific Proposal Grades

	Male	Female	Difference
poor	0.021	0.026	-0.006
average	0.145	0.176	-0.031
good	0.320	0.344	-0.025
very good	0.320	0.294	0.026
excellent	0.171	0.141	0.031
outstanding	0.023	0.018	0.005

This conclusion is reinforced when we look at the plots above, which show the probability and cumulative probability curves of getting each grade for male and female applicants. They follow more or less the same trend and the only difference, as we've seen from the table above, is that women are slightly more likely to get a "very good" rather than an "excellent". We therefore conclude that gender is a significant, but a not necessarily practically relevant, predictor of the different grades.

(2) **Applicant Track assessment**: The next model we used has ApplicantTrack as a response variable and the following predictors: Gender, Division, PercentFemale, IsContinuation, InstType, log(AmountRequested), Semester, the interaction between Gender and Division and the interaction between Division and PercentFemale. Again we fitted the same model without Gender and compared it to the one with gender using the likelihood ratio test included in the anova() function. We got a p-value of 2.7×10^{-4}, meaning that for the grades given to the main applicant track record, gender should be included in the model. We see from the confidence interval plot that the interaction Gender:Division is significant. The other significant variables seem to be: Division, Institution type, IsContinuation and the percentage of female reviewers. Finally, the interaction between Division and the percentage of female seems to have some importance too. However, once again, gender is significant but not practically relevant, as evidenced by the negligible differences between genders in the table below.

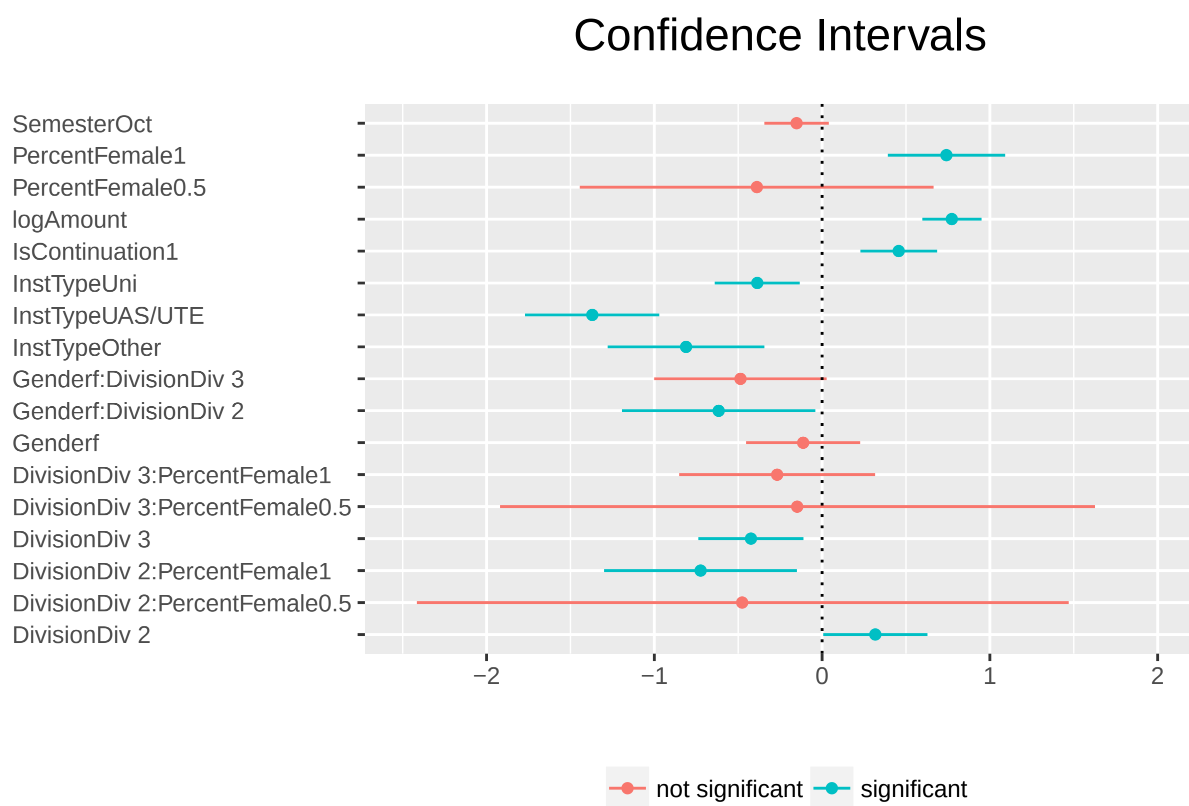
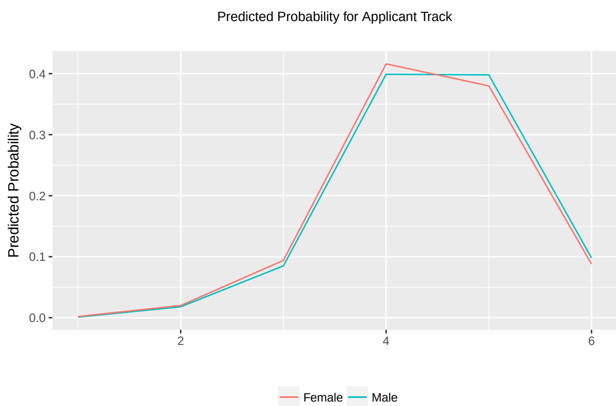
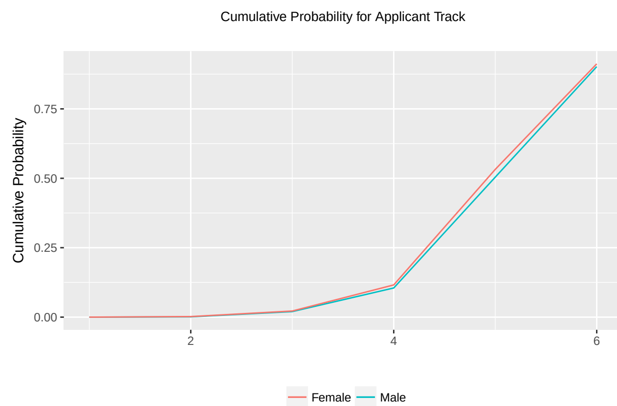


Table 6: Predicted probabilities of getting different Applicant Track grades

	Male	Female	Difference
poor	0.001	0.002	0.000
average	0.018	0.020	-0.002
good	0.085	0.094	-0.009
very good	0.399	0.416	-0.017
excellent	0.398	0.380	0.019
outstanding	0.098	0.088	0.010

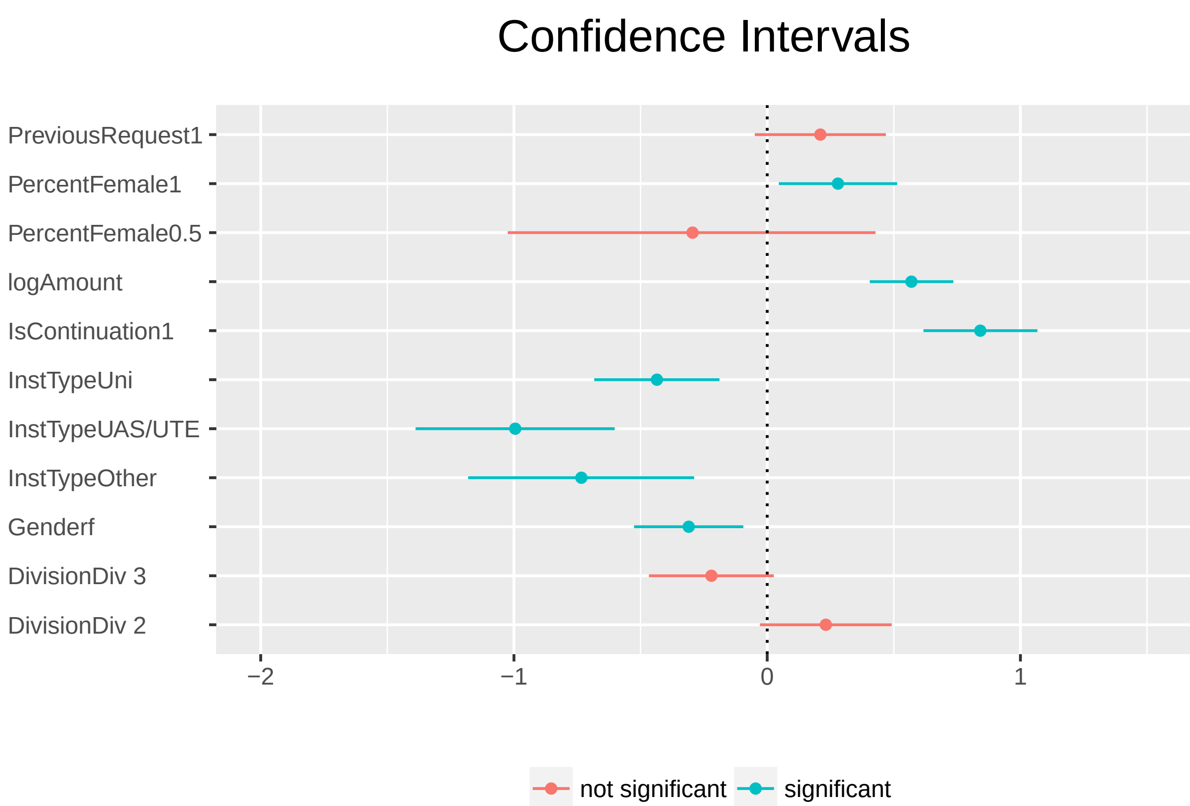
We computed the predicted probability of getting each grade for both male and female and the difference between the two. The average difference of the cumulative probability is here 0.0095, very close to zero. From the plot below we see that there is almost no difference between women and men probabilities.





(3) **Ranking**: This last model has Ranking as a response and Gender, Division, PercentFemale, IsContinuation, InstType, PreviousRequest and logAmount as predictors. We are not considering here the grades given to the applicant track record and to the project, as we just want to see the influence of the demographic data and the project information in each grade. A comparison of this model with the same one without gender may suggest that gender is significant to the model: p-value of 0.0049.

We can also see from the plot below that the upper bound of the confidence interval of gender is not so far from 0. This also suggests that the effect may not be so large.

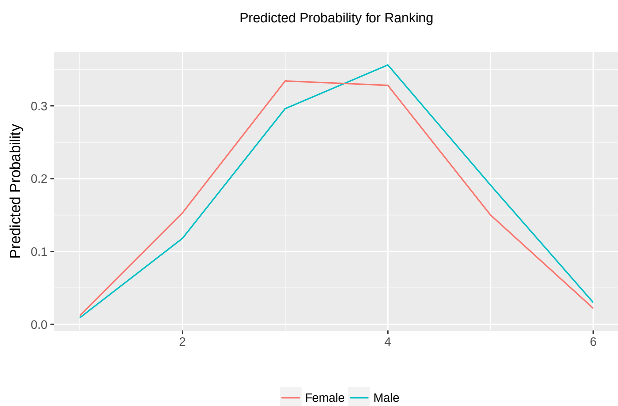


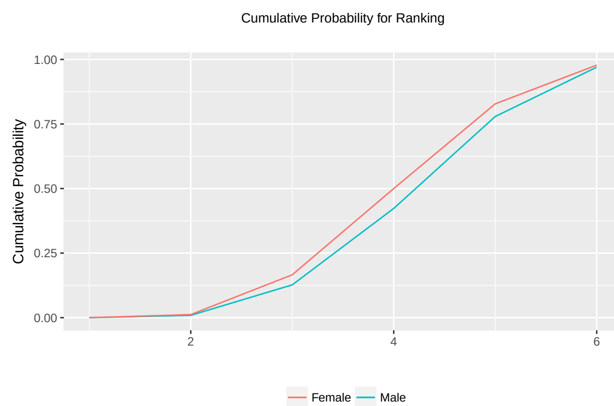
Once again the differences between male and female predicted probabilities are not too large, so we do not find gender to be a meaningful predictor of the grade. However, we do notice that the cumulative difference between male and female for the top three grades (A, AB, B) is 7.7pp. While this is still a small difference, it is worth continuing to pay attention to since the aggregate effect is more noticeable than the individual

Table 7: Predicted probabilities of getting different Ranking grades

	Male	Female	Difference
D	0.009	0.012	-0.003
C	0.118	0.153	-0.035
BC	0.296	0.334	-0.038
B	0.356	0.328	0.028
AB	0.191	0.150	0.041
A	0.030	0.022	0.008

differences.





Results

Gender Bias in the Funding Decision

When assessing the assocation between Gender and IsApproved in each of the three Divisions at the SNSF, we failed to reject the null hypothesis that the two variables are independent at a 5% significance level. Thus we failed to find evidence at an overarching level of gender bias in the funding decision.

External Review Step

Within the external review step, we discovered that a logistic regression using data from the external step is not a good explanation of the variation in the approval of applications (around 42%). We nevertheless look into the influence of gender at this stage, but couldn't find evidence of its effect on the final decision.

When looking at the different grades in this step, we found that gender did not have a significant influence on ProposalCombined or OverallGrade. We did find that gender was a significant predictor for the grade of Applicant Track, but the impact in terms of the difference in predictive probabilities was rather low.

Internal Review Step

The logistic regression using the data from the internal review step is a much better explanation of the variation in the approval of applications (around 70% of the variability is explained). Gender was not a significant predictor in this regression.

When looking the predictors of the different grades, we found that gender was a significant predictor for the ProjectAsssessment and Ranking grades, but not for the ApplicantTrack. The individual effects were small (in terms of the difference in predicted probabilities), so it's effect was significant but not practically relevant. The combined difference in predictive probabilities of top grades given to male vs. female applicants was more notable than the individual grades, albiet the overall impact was still reletively small.

Considering all the analysis that we have done so far, we conclude that there is no evidence of gender bias in the funding decision at the Swiss National Science Foundation. Gender is generally not an important variable in the regressions we performed, and where it was a significant predictor, the difference in the predicted probability between male and female applicants is not practically relevant.

Relative Importance of the Different Steps

Our second research question was to assess the relative importance of each step in the process, and the relative importance of each criteria within each step. To answer this question, we took the following approach:

(3) **Most Important Step: Logistic Regression:** We first fit a logistic regression with IsApproved as our binary response variable using the glm function, with the demographic data and the summary grades from the external and internal step to assess which step of the process was most important to determining the final funding decision.

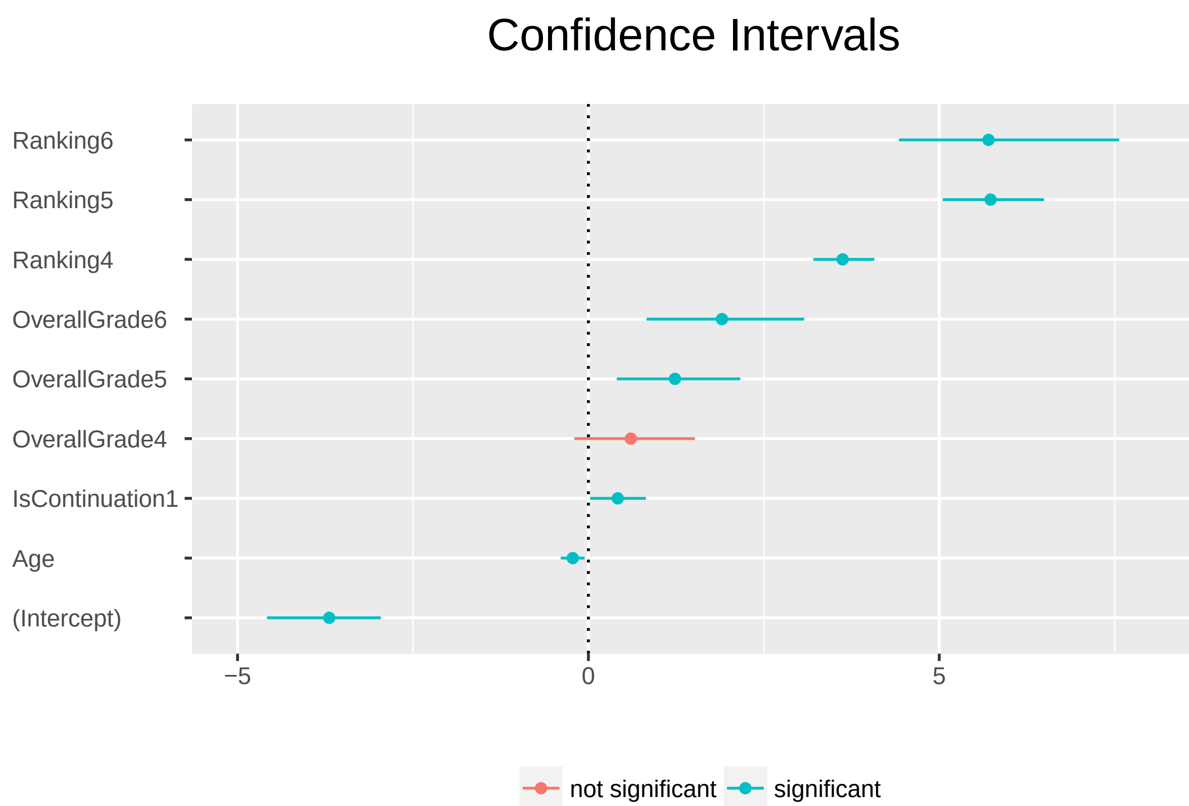
(4) **Most Important Criteria within Each Step:** We then assessed which criteria was most important within each step to predict the overall grade given to an application (OverallGrade in the External step, Ranking in the Internal step). We did this by fitting an Ordinal Regression (with the OverallGrade for external and Ranking for internal as the response) with standardized coefficients, and then comparing the magnitude of each predictor. We use this approach for first the external and then the internal step.

Most Important Step: Logistic Regression

To approach the question of which step in the process is most important, we first fit a logistic regression with IsApproved as our binary response variable using the glm function in R. We fit a full model with all potential demographic predictors and interactions across the different steps of the process, and the summary grade given to an application in the external (OverallGrade) and internal (Ranking) step. To address the first part of the question (relative importance of each step in the process), we used only the summary grade in each step due to the correlation between the individual grades given within each step and the summary grade given. With the full model (predictors: Gender, Division, Age, IsContinuation, InstType, log(AmountRequested), PercentFemale, Ranking, OverallGrade, Gender:Division, PercentFemale:Gender), we achieved a pseudo-R^2 value of 0.7251, indicating that percent of the variation in IsApproved can be explained by the model.

As our goal was to explain the most important factors, we then did backwards variable selection using the AIC. This left us with a model with only 4 predictors: Ranking, OverallGrade, Age, and IsContinuation. The pseudo-R^2 measure of this model is 0.7234, which indicates that this simplified model nearly explains exactly as much variance in the data as the full model, and so we can be content to use just the small model.

Now that we've reduced our model to 4 predictors, we wanted to understand exactly how important each of those predictors are to the final funding decision. To do this, we looked at the confidence intervals of the coefficients to see which had the largest impact. To do this, we needed to first standardize our continuous variables (Age). When we plotted the confidence intervals of our coefficients, we can see that the internal Ranking has by far the largest coefficient, and thus the biggest impact on the final funding decision.



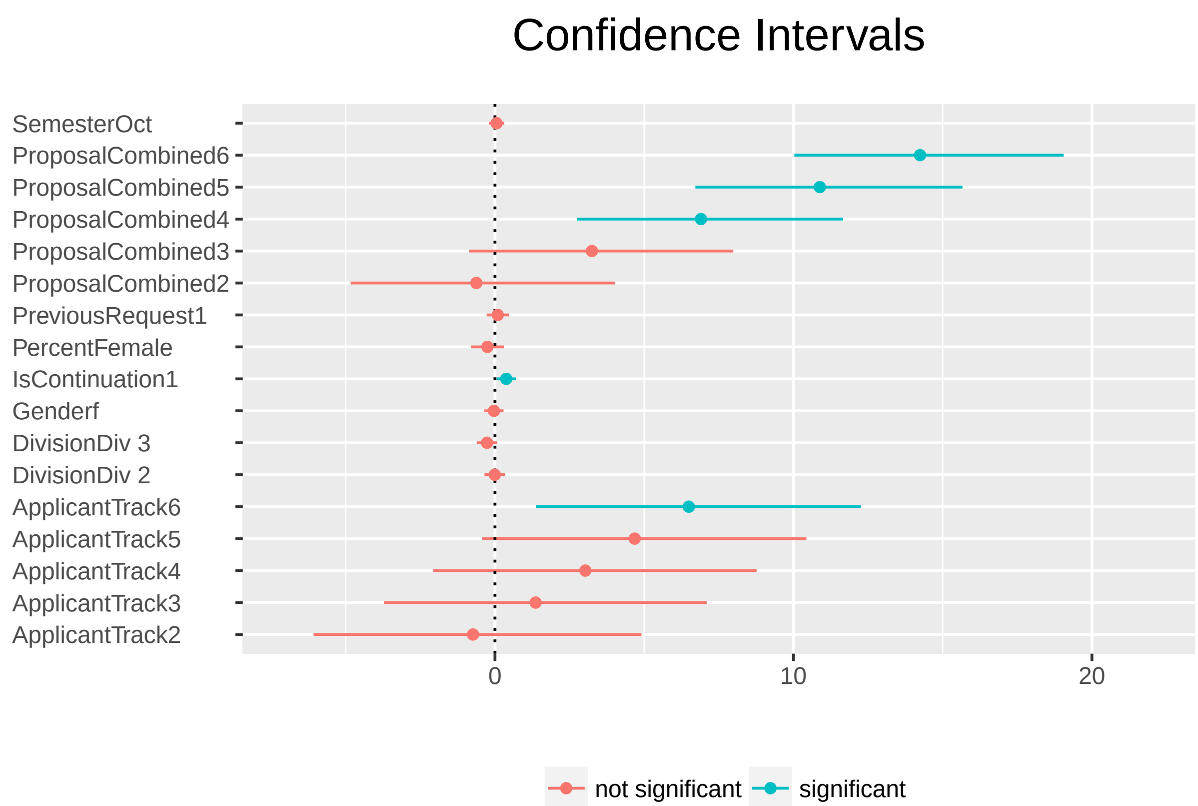
Most Important Criteria Within Each Step

The second aspect of this question was to identify what was the most important criteria within each step.

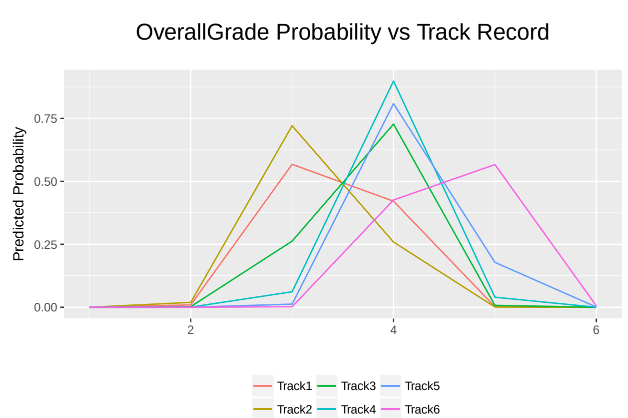
External Review Step

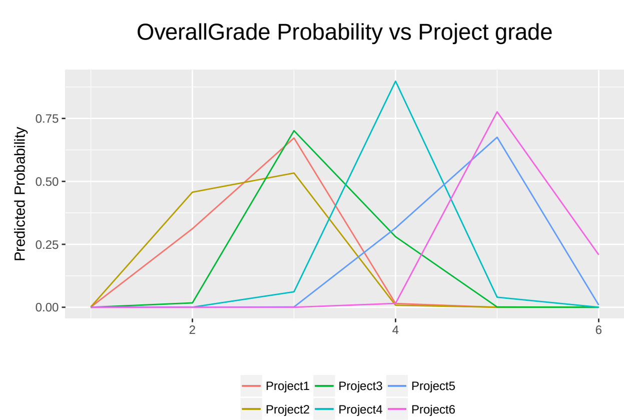
In order to check which variable has the biggest influence on the final grade given by the external reviewers, we fitted an ordinal regression using the "OverallGrade" grade as the multinomial response and ApplicantTrack, ProposalCombined grade and demographic data as predictors. We started as always fitting the full model and then did variable selection using the AIC. The explanatory variables included in the final model are: ApplicantTrack, ProposalCombined, Gender, Division, PercentFemale, IsContinuation, PreviousRequest and Semester.

We computed the predicted probabilities for each of the Overall grade, varying the Applicant Track grade (plot on the left) and the ProposalCombined grade (plot on the right). It is clear that the same variation in grade implies different changes in probabilities: when the project grade changes, the probability variation is much bigger. The quality of the project has a greater influence on the OverallGrade, compared to the track record.



From the plot above we can see that the variable whose coefficients are at the biggest distance from 0 is the ProposalCombined, when the grade is 5 or 6. The grades given to the track record of the applicant are significant too, since the confidence intervals don't include 0 but we can see that the effect is smaller.



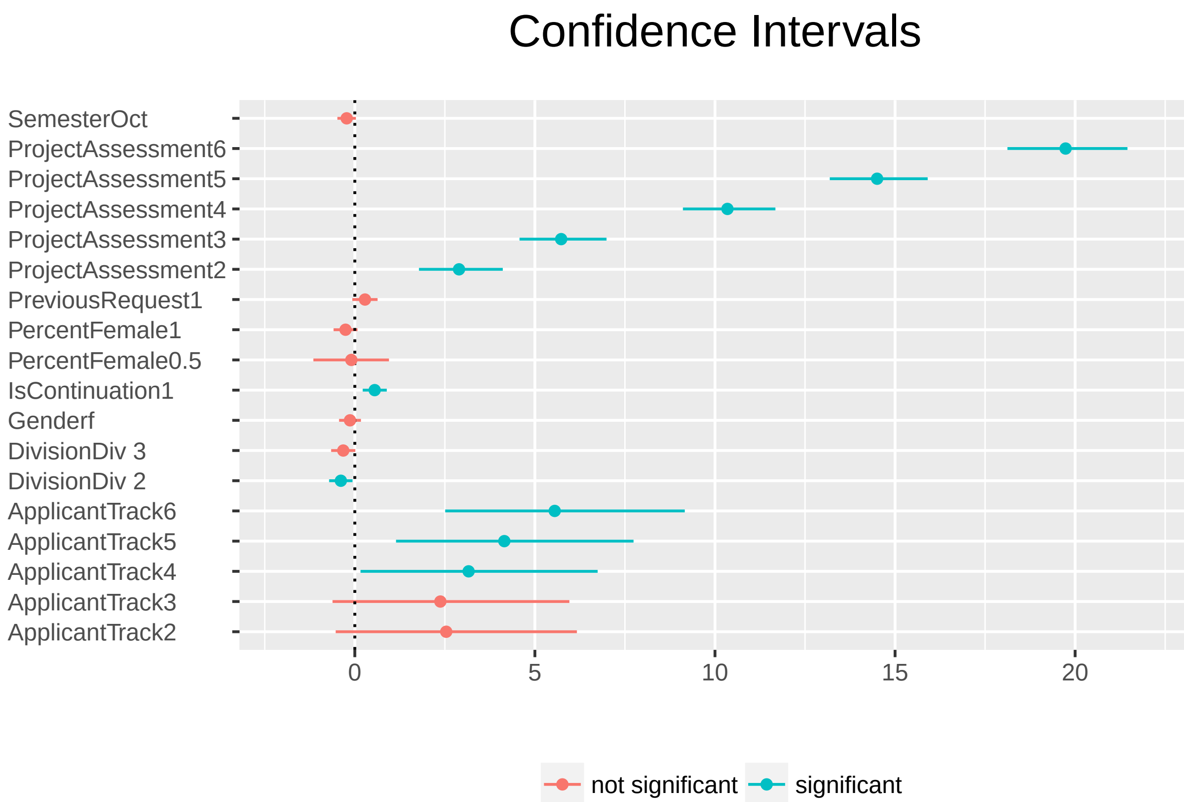


We also computed the predicted probabilities for each of the OverallGrade grades, varying the Applicant Track grade (plot on the left) and the ProposalCombined grade (plot on the right). It is clear that the same variation in grade implies different changes in probabilities: when the project grade changes, the probability variation is much bigger. Notably, in the right graph, the highest probability to the final grade corresponds with the grade given to the ProposalCombined for grades 3, 4, 5 and 6. On the left, receiving a 3, 4 or 5 for the ApplicantTrack record all correspond to most likely receiving an overall grade of 4. This confirms what we found before: the quality of the project has a greater influence on the final Overall Grade, compared to the track record.

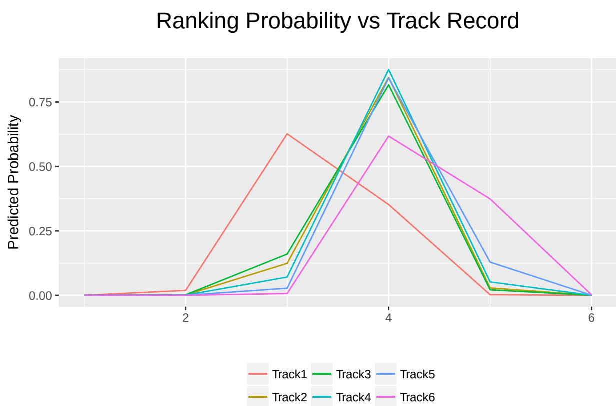
Internal Review Step

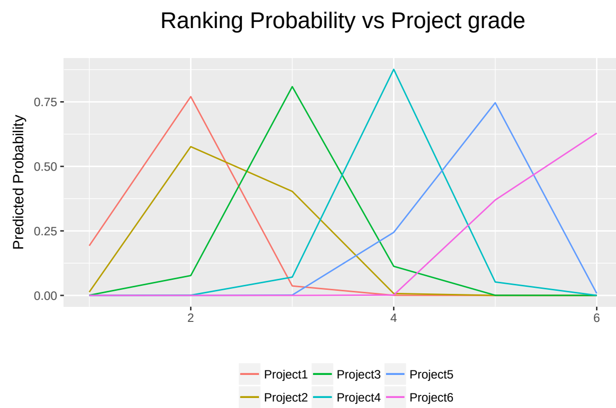
In order to check which variable has the biggest influence on the final grade given by the internal referees, we fitted an ordinal regression using the Ranking grade as a multinomial response and ApplicantTrack, ProjectAssessment and demographic data as predictors. We started as always fitting the full model and did variable selection using the AIC. The explanatory variables included in the final model are: ApplicantTrack, ProjectAssessment, Gender, Division, PercentFemale, IsContinuation, PreviousRequest and Semester.

We also computed the predicted probabilities for each of the Ranking grades, varying the ApplicantTrack grade (plot on the left) and the ProjecAssessment grade (plot on the right). It is clear that the same variation in grade implies different changes in probabilities: when the project grade changes, the probability variation is much bigger. This confirms what we found before: the quality of the project has a greater influence on the final Ranking, compared to the track record.



From the plot above we can see that the variable whose confidence intervals are at the biggest distance from 0 is the Project Assessment. The grades given to the track record of the applicant are significant too, since the confidence intervals don't include 0 but we can see that the effect is smaller.





We also computed the predicted probabilities for each of the Ranking grade, varying the Applicant Track grade (plot on the left) and the Project Assessment grade (plot on the right). It is clear that the same variation in grade implies different changes in probabilities: when the project grade changes, the probability variation is much bigger. This confirms what we found before: the quality of the project has a greater influence on the final ranking, compared to the track record.

Conclusion

Gender bias not at an overall level. Within the external step, some small differences in the grades given to women, in particular, fewer high grades given to women. The difference is small, and so overall still not much evidence of a gender bias.

Futher Analysis

Appendix

A.1 Detailed Data Description

Applications

- **AmountRequested**: Rounded to the next 10k CHF
- **AmountGranted**: Rounded to the next 10k CHF
- **IsApproved**: 1 if the application was approved, 0 if it was rejected
- **GradeFinal**: Comparative ranking of the application as determined by the evaluation body (the division of the National Research Council). A: “belongs to the 10% best percent”; AB: “10% are worse, 75% are better”; B: “50% are worse, 25% are better”; BC: " 25% are worse, 50% are better“; C”10% are worse, 75% are better“; D:”90% of the applications are better"
- **Division**: Evaluation Body in which the application was evaluated. Division 1 evaluates Social Sciences and Humanities; Division 2 Mathematics, Natural Sciences and Engineering; Division 3 Biology and Medicine
- **MainDiscipline**: as chosen by the applicant from the SNSF discipline list
- **MainDisciplineLevel2**: category in the SNF discipline list grouping disciplines into fields of research
- **CallTitle**: Call for proposals under which the application was submitted. Applications from the same Call are evaluated together, i.e. in competition to each other
- **CallEndDate**: Submission deadline of the Call
- **ResponsibleApplicantAcademicAgeAtSubmission**: Years since the applicant’s PhD at time of submission; data only available since mid 2016
- **ResponsibleApplicantAgeAtSubmission**: Biological age of the applicant at time of submission; data only available since mid 2016
- **ResponsibleApplicantProfessorshipType**: employment situation of the applicant at time of submission; data only available since mid 2016
- **Gender**: of the main applicant
- **NationalityIsoCode**: Nationality of the main applicant
- **IsHasPreviousProjectRequested**: 0 if it is the applicant’s first application at the SNSF, 1 if not
- **InstType**: Type of institution where the applicant is employed
- **IsContinuation**: 1 if the project is a thematic continuation of a previously approved project, 0 if not
- **ProjectID**: Anonymized identifier of the application

Referee Grades

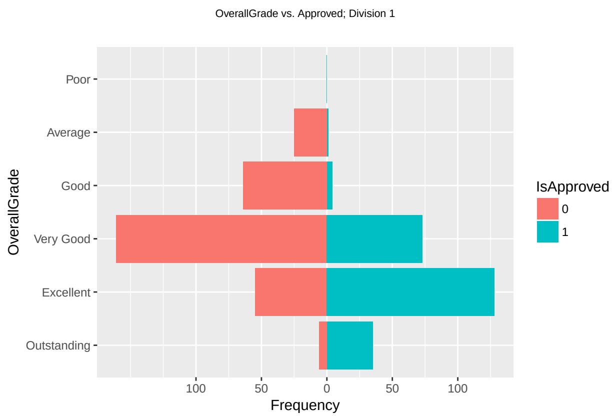
- **Question**: Evaluation criterion
- **QuestionRating**: The (co-)referee’s assessment of the evaluation criterion
- **OverallRanking**: The (co-)referee’s overall comparative ranking of the application. A: “belongs to the 10% best percent”; same scale as the GradeFinal
- **RefereeRole**: Some applications have one referee evaluation, some have two. The role indicates who was the primary and who was the secondary referee (also called co-referee)
- **RefereeGender**
- **IDs**: Anonymized identifiers of the application, the referee and the evaluation by the referee

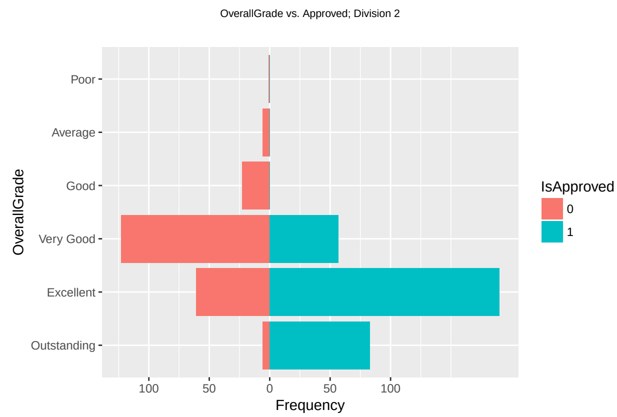
Reviews

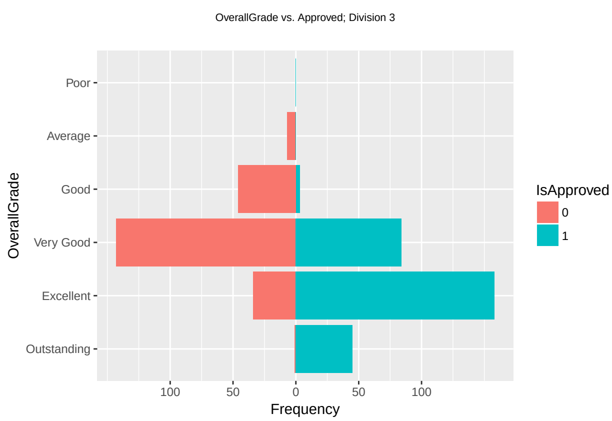
- **Question**: Evaluation criterion
- **QuestionRating**: The external reviewer's assessment of the evaluation criterion
- **OverallGrade**: The external reviewer's overall assessment of the application
- **SourcePerson**: Who suggested the reviewer?
- **Gender**
- **Country**: where the reviewer is located. Not always known
- **EmailEnding**: ending of the reviewer's email address. Might be used as an approximation of the country where the reviewer is located in cases where this data is missing
- **IDs**: Anonymized identifiers of the application, the reviewer and the review

A.2 Exploratory Analysis

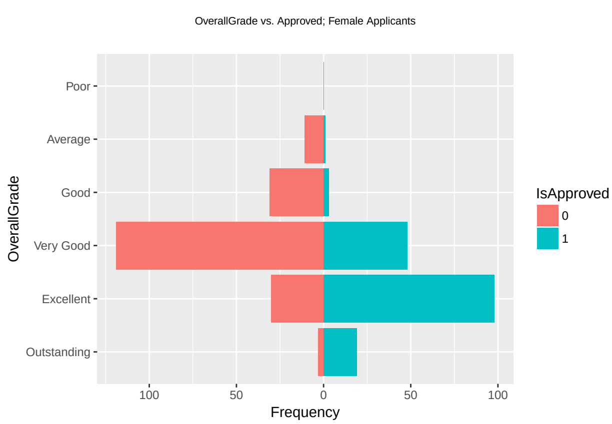
OverallGrade vs. IsApproved, by Division

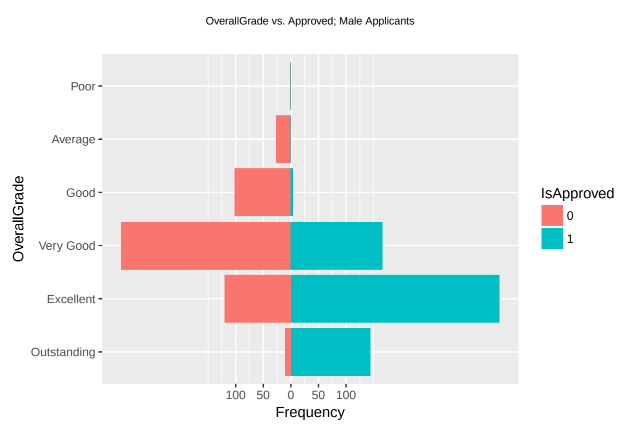






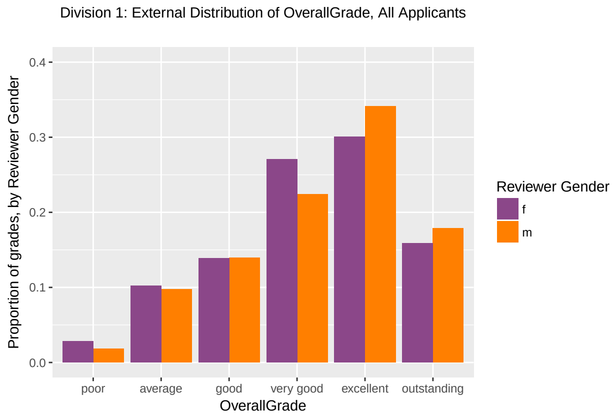
OverallGrade vs. IsApproved, by Gender

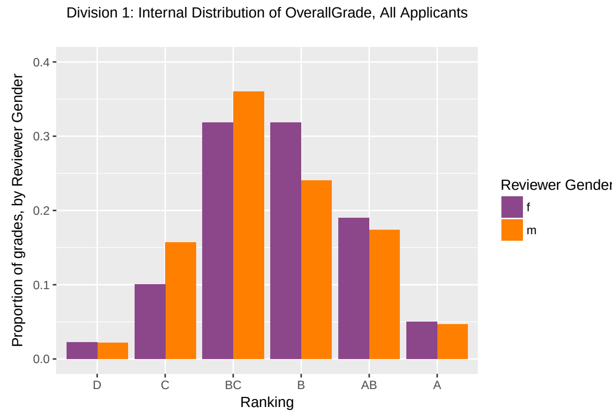




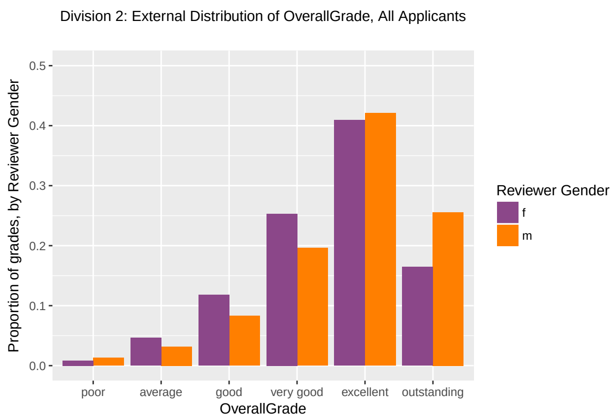
Distribution of Grades by Reviewer Gender & Division (All Applicants)

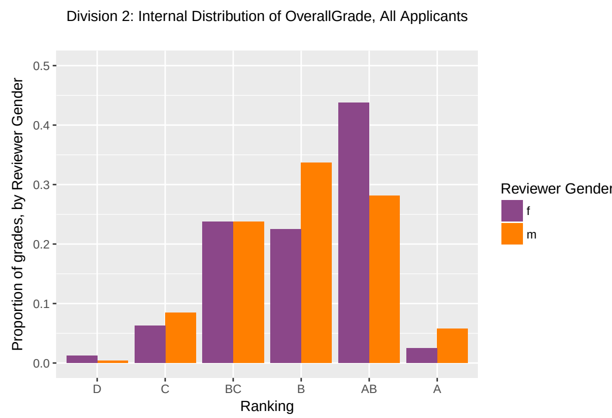
Division 1



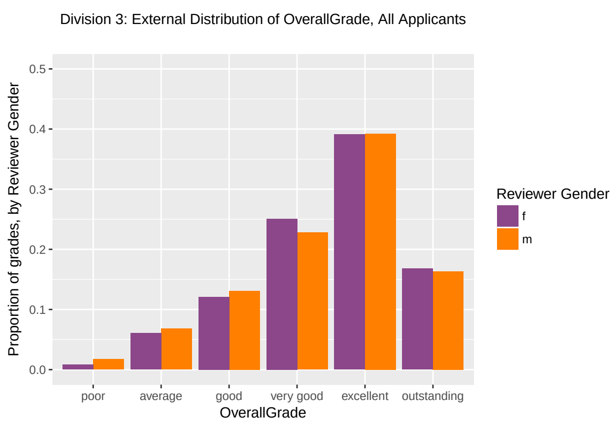


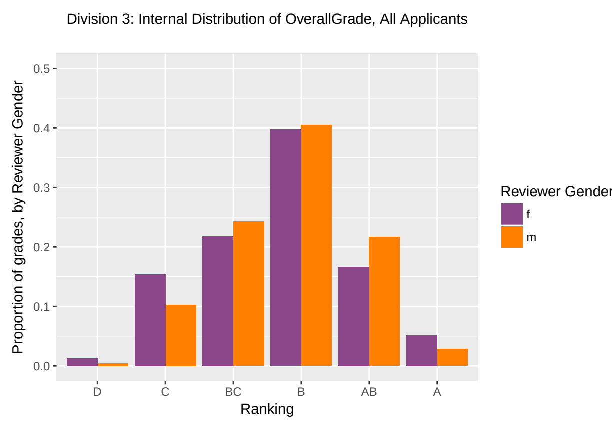
Division 2





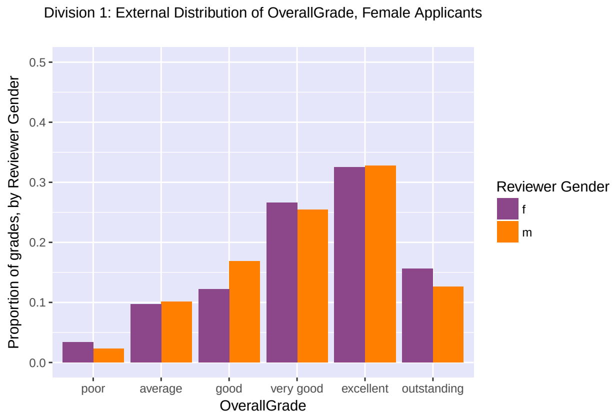
Division 3

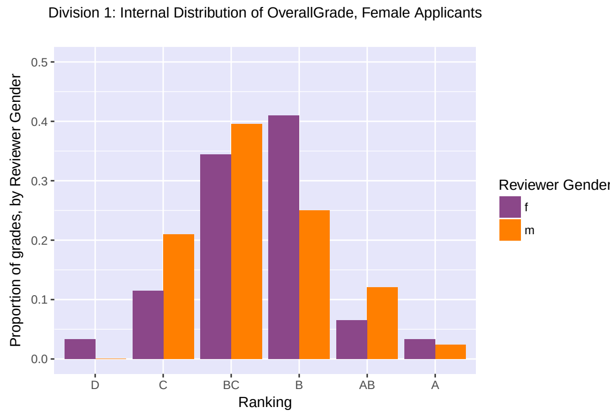




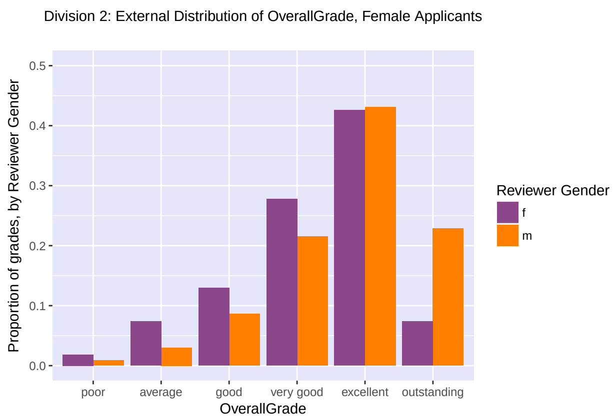
Distribution of Grades by Reviewer Gender & Division (Female Applicants)

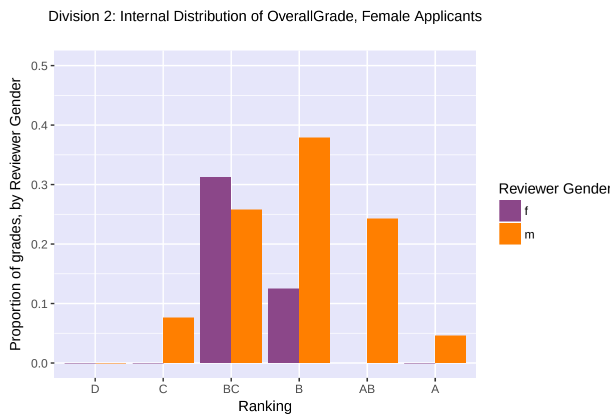
Division 1



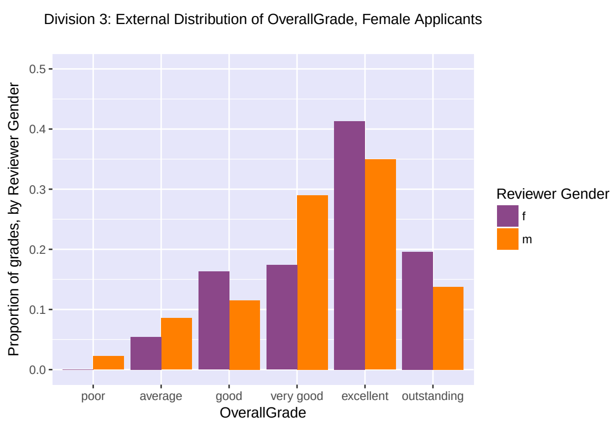


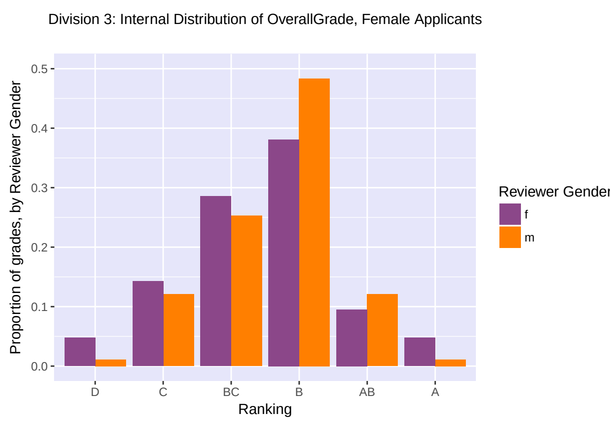
Division 2





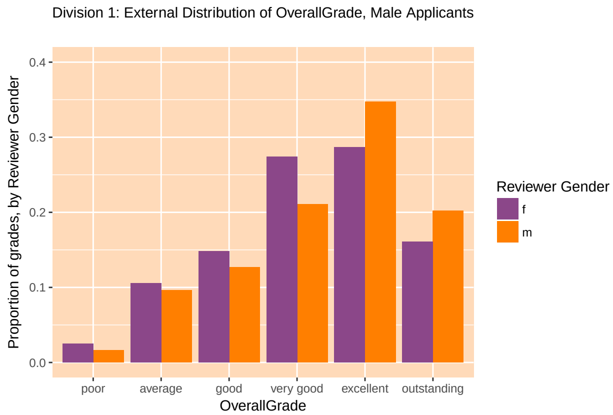
Division 3

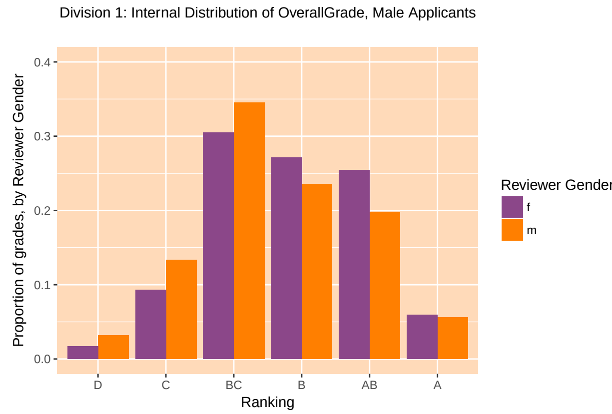




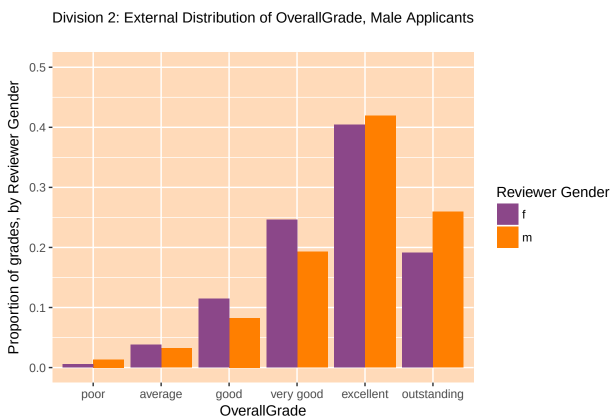
Distribution of Grades by Reviewer Gender & Division (Male Applicants)

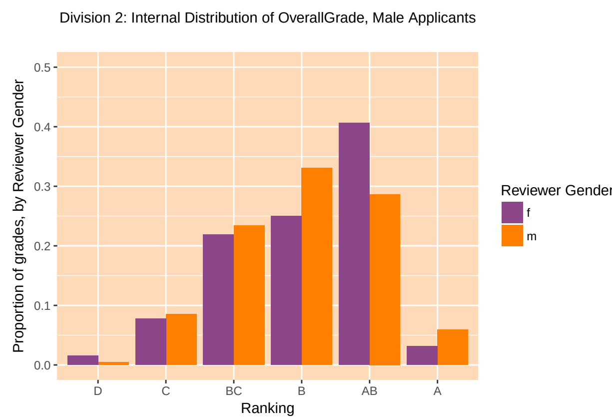
Division 1



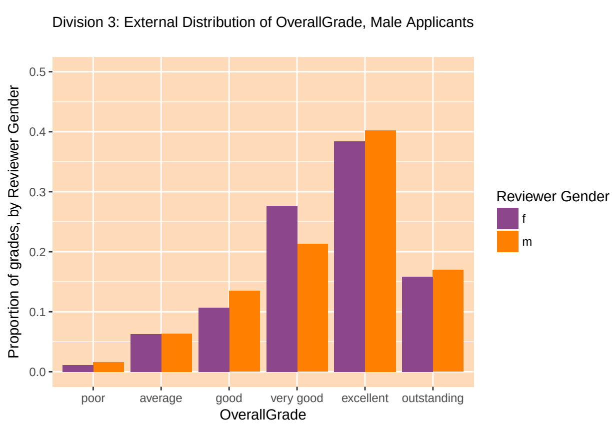


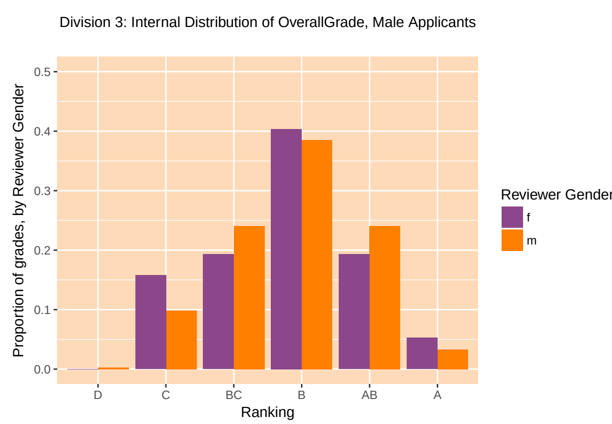
Division 2





Division 3





A.3 Introduction to Cumulative Link Models (CLM)

A cumulative link model is a model for an ordinal response variable Y_i that can fall in $j = 1, ..., J$ categories. So Y_i follows a multinomial distribution with parameter $\boldsymbol{\pi}$, where π_{ij} denotes the probability that the i-th observation falls into response category j. We denote the cumulative probabilities as

$$\gamma_{ij} = P(Y_i \leq j) = \pi_{i1} + ... + \pi_{ij}$$

. Then we consider the logit function as link function: the cumulative logits are defined as

$$logit(\gamma_{ij}) = logit(P(Y_i \leq j)) = log\frac{P(Y_i \leq j)}{1 - P(Y_i \leq j)}$$

.

The *cumulative logit model* is a regression model for cumulative logits and it can also be written as: $logit(\gamma_{ij}) = \theta_j - {x_i}^T\beta$, where x_i is a vector of explanatory variables for the i-th observation and β is the corresponding set of regression parameters.

This means that each cumulative logit for each j has its own intercept, given by the parameter θ_j. The regression part ${x_i}^T$ is independent of j and so the effect β of the explanatory variable is the same for each of the $J - 1$ cumulative logits.

The interpretation of the model is the following: the log-odds of two cumulative distributions measure how likely the response is to be in category j or below versus falling in a category higher than j.

Cumulative logit models can be fitted with the function `clm` from package `ordinal`. The arguments used are the formula (`response    covariates`), the dataset and the link function used. By default `clm` uses the logit link. The `summary` of the clm object provides basic information about the fit of the model: coefficient tables for the regression variables and for the cut-points, Wald tests and corresponding p-values.

A.4 External Step Analysis

Logistic Regression final model

```
Call:
glm(formula = IsApproved ~ ApplicantTrack + ScientificRelevance +
    Suitability + PercentFemale + Age + Gender + Division + IsContinuation +
    InstType + Semester + Division:InstType, family = "binomial",
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3819  -0.8248   0.3575   0.7624   2.4032

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -3.029987   0.761163  -3.981 6.87e-05 ***
ApplicantTrack4                0.969484   0.503628   1.925 0.054229 .
ApplicantTrack5                1.392174   0.510720   2.726 0.006412 **
ApplicantTrack6                2.016852   0.537235   3.754 0.000174 ***
ScientificRelevance4           0.986281   0.391741   2.518 0.011813 *
ScientificRelevance5           1.631665   0.406202   4.017 5.90e-05 ***
ScientificRelevance6           1.776094   0.464162   3.826 0.000130 ***
Suitability4                   0.657609   0.215775   3.048 0.002306 **
Suitability5                   1.781169   0.250140   7.121 1.07e-12 ***
Suitability6                   1.986616   0.390191   5.091 3.55e-07 ***
PercentFemale                 -0.418836   0.261696  -1.600 0.109494
Age                           -0.004531   0.007821  -0.579 0.562333
Genderf                       -0.112332   0.150209  -0.748 0.454557
DivisionDiv 2                 -0.516975   0.431020  -1.199 0.230364
DivisionDiv 3                  0.387478   0.543470   0.713 0.475864
IsContinuation1                0.720271   0.158877   4.534 5.80e-06 ***
InstTypeOther                 -0.174505   0.615954  -0.283 0.776940
InstTypeUAS/UTE               -0.159980   0.464778  -0.344 0.730690
InstTypeUni                   -0.472194   0.423065  -1.116 0.264368
SemesterOktober                0.183584   0.123977   1.481 0.138664
DivisionDiv 2:InstTypeOther    0.321877   0.807478   0.399 0.690174
DivisionDiv 3:InstTypeOther   -0.743797   0.841546  -0.884 0.376779
DivisionDiv 2:InstTypeUAS/UTE -1.416883   0.812385  -1.744 0.081141 .
DivisionDiv 3:InstTypeUAS/UTE -0.191572   0.987872  -0.194 0.846235
DivisionDiv 2:InstTypeUni      0.517898   0.479246   1.081 0.279853
DivisionDiv 3:InstTypeUni     -0.279593   0.569949  -0.491 0.623740
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2243.9  on 1622  degrees of freedom
Residual deviance: 1630.2  on 1597  degrees of freedom
AIC: 1682.2

Number of Fisher Scoring iterations: 5
```

Ordinal Regression for the Scientific Proposal

Summary of the final model:

```
formula:
ProposalCombined ~ Gender + Division + PercentFemale + IsContinuation + InstType + logAmount
data:    external_regression_data

 link  threshold nobs logLik   AIC     niter max.grad cond.H
 logit flexible  1623 -1949.71 3927.42 7(0)  8.06e-10 4.9e+05

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
Genderf         -0.18423    0.11413  -1.614  0.10649
DivisionDiv 2    0.36126    0.14190   2.546  0.01090 *
DivisionDiv 3   -0.28708    0.13153  -2.183  0.02907 *
PercentFemale   -0.43375    0.20023  -2.166  0.03029 *
IsContinuation1  0.57973    0.11661   4.971 6.65e-07 ***
InstTypeOther   -0.63461    0.23714  -2.676  0.00745 **
InstTypeUAS/UTE -0.83721    0.20535  -4.077 4.56e-05 ***
InstTypeUni     -0.27226    0.13099  -2.078  0.03767 *
logAmount        0.44651    0.08627   5.176 2.27e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2   -1.280      1.310  -0.977
2|3    1.798      1.114   1.613
3|4    3.416      1.108   3.084
4|5    5.709      1.112   5.132
5|6    8.174      1.123   7.276
```

Odds Ratios and Confidence intervals:

Table 8: OR and corresponding Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.83	0.66	1.04
DivisionDiv 2	1.44	1.09	1.90
DivisionDiv 3	0.75	0.58	0.97
PercentFemale	0.65	0.44	0.96
IsContinuation1	1.79	1.42	2.25
InstTypeOther	0.53	0.33	0.84
InstTypeUAS/UTE	0.43	0.29	0.65
InstTypeUni	0.76	0.59	0.98
logAmount	1.56	1.32	1.85

Ordinal Regression for the Applicant Track

Summary of the final model:

```
formula:
ApplicantTrack ~ Gender + Division + PercentFemale + IsContinuation + InstType + logAmount + Gender
data:    external_regression_data

 link  threshold nobs logLik   AIC     niter
 logit flexible  1623 -1895.84 3821.69 7(0)
 max.grad cond.H
 7.99e-12 5.1e+05

Coefficients:
                      Estimate Std. Error z value
Genderf               -0.51334    0.16102  -3.188
DivisionDiv 2          0.51317    0.14066   3.648
DivisionDiv 3         -0.19141    0.12973  -1.475
PercentFemale         -0.79751    0.23535  -3.389
IsContinuation1        0.47308    0.11730   4.033
InstTypeOther         -0.55931    0.23642  -2.366
InstTypeUAS/UTE       -1.26531    0.20288  -6.237
InstTypeUni           -0.23820    0.13069  -1.823
logAmount              0.55195    0.08643   6.386
Genderf:PercentFemale  1.04403    0.41418   2.521
                      Pr(>|z|)
Genderf               0.001433 **
DivisionDiv 2         0.000264 ***
DivisionDiv 3         0.140095
PercentFemale         0.000703 ***
IsContinuation1       5.51e-05 ***
InstTypeOther         0.017996 *
InstTypeUAS/UTE       4.46e-10 ***
InstTypeUni           0.068343 .
logAmount             1.70e-10 ***
Genderf:PercentFemale 0.011711 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2  -0.7523     1.4897  -0.505
2|3   2.0350     1.1325   1.797
3|4   3.8149     1.1106   3.435
4|5   6.1707     1.1115   5.551
5|6   8.2987     1.1208   7.404
```

Odds Ratios and Confidence intervals:

Table 9: OR and corresponding Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.60	0.44	0.82
DivisionDiv 2	1.67	1.27	2.20
DivisionDiv 3	0.83	0.64	1.06
PercentFemale	0.45	0.28	0.71
IsContinuation1	1.60	1.28	2.02
InstTypeOther	0.57	0.36	0.91
InstTypeUAS/UTE	0.28	0.19	0.42
InstTypeUni	0.79	0.61	1.02
logAmount	1.74	1.47	2.06
Genderf:PercentFemale	2.84	1.26	6.40

Ordinal Regression for the Overall Grade

Summary of the final model:

```
formula:
OverallGrade ~ Gender + PercentFemale + Division + IsContinuation + PreviousRequest + InstType + log
data:    external_regression_data

 link  threshold nobs logLik   AIC     niter
 logit flexible  1623 -1989.35 4008.69 8(0)
 max.grad cond.H
 3.51e-12 5.2e+05

Coefficients:
                 Estimate Std. Error z value
Genderf          -0.18044    0.11418  -1.580
PercentFemale    -0.52959    0.20014  -2.646
DivisionDiv 2     0.38571    0.14093   2.737
DivisionDiv 3    -0.33079    0.13064  -2.532
IsContinuation1   0.65198    0.11936   5.462
PreviousRequest1  0.06592    0.13524   0.487
InstTypeOther    -0.50460    0.23624  -2.136
InstTypeUAS/UTE  -0.85837    0.20480  -4.191
InstTypeUni      -0.23708    0.13043  -1.818
logAmount         0.44998    0.08581   5.244
                 Pr(>|z|)
Genderf           0.11405
PercentFemale     0.00814 **
DivisionDiv 2     0.00620 **
DivisionDiv 3     0.01134 *
IsContinuation1  4.70e-08 ***
PreviousRequest1  0.62593
InstTypeOther     0.03268 *
InstTypeUAS/UTE  2.77e-05 ***
InstTypeUni       0.06912 .
logAmount        1.57e-07 ***
---
Signif. codes:
```

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2   -1.877      1.492  -1.257
2|3    1.845      1.120   1.648
3|4    3.478      1.113   3.126
4|5    5.727      1.117   5.129
5|6    7.971      1.126   7.077
```

Odds Ratios and Confidence intervals:

Table 10: OR and corresponding COnfidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.83	0.67	1.04
PercentFemale	0.59	0.40	0.87
DivisionDiv 2	1.47	1.12	1.94
DivisionDiv 3	0.72	0.56	0.93
IsContinuation1	1.92	1.52	2.43
PreviousRequest1	1.07	0.82	1.39
InstTypeOther	0.60	0.38	0.96
InstTypeUAS/UTE	0.42	0.28	0.63
InstTypeUni	0.79	0.61	1.02
logAmount	1.57	1.33	1.86

A.4 Internal Step Analysis

Logistic Regression

Summary of the final model

```
Call:
glm(formula = IsApproved ~ Gender + Age + Semester + IsContinuation +
    PercentFemale + ApplicantTrack + ProjectAssessment + logAmount,
    family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6763  -0.4069   0.2040   0.5752   2.6648

Coefficients:
                   Estimate Std. Error z value
(Intercept)        -0.55687    2.05505  -0.271
Genderf             0.14332    0.19051   0.752
Age                -0.01377    0.01017  -1.354
SemesterOct         0.19430    0.16517   1.176
IsContinuation1     0.58642    0.19803   2.961
PercentFemale      -0.38897    0.20067  -1.938
ApplicantTrack4     0.80144    0.29538   2.713
ApplicantTrack5     1.25919    0.30685   4.104
ApplicantTrack6     1.16908    0.42310   2.763
ProjectAssessment4  3.24147    0.18380  17.636
ProjectAssessment5  5.46123    0.32697  16.703
ProjectAssessment6  5.66171    0.77317   7.323
logAmount          -0.17391    0.15548  -1.118
                   Pr(>|z|)
(Intercept)         0.78641
Genderf             0.45187
Age                 0.17565
SemesterOct         0.23943
IsContinuation1     0.00306 **
PercentFemale       0.05258 .
ApplicantTrack4     0.00666 **
ApplicantTrack5    4.07e-05 ***
ApplicantTrack6     0.00572 **
ProjectAssessment4  < 2e-16 ***
ProjectAssessment5  < 2e-16 ***
ProjectAssessment6 2.43e-13 ***
logAmount           0.26335
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2243.9  on 1622  degrees of freedom
Residual deviance: 1073.2  on 1610  degrees of freedom
AIC: 1099.2
```

```
Number of Fisher Scoring iterations: 6
```

Ordinal Regression for the Scientific Proposal

Summary of the final model:

```
    formula:
    ProjectAssessment ~ Gender + Division + PercentFemale + Age + IsContinuation + InstType + logAmount
    data:    internal_regression_data

     link  threshold nobs logLik   AIC     niter
     logit flexible  1623 -2351.71 4735.43 6(0)
     max.grad cond.H
     3.09e-09 7.9e+06

    Coefficients:
                      Estimate Std. Error z value
    Genderf          -0.247430   0.110411  -2.241
    DivisionDiv 2     0.469429   0.133517   3.516
    DivisionDiv 3     0.005880   0.125754   0.047
    PercentFemale0.5 -0.323785   0.374886  -0.864
    PercentFemale1    0.342865   0.118669   2.889
    Age               0.008575   0.005778   1.484
    IsContinuation1   0.735298   0.116026   6.337
    InstTypeOther    -0.612199   0.227930  -2.686
    InstTypeUAS/UTE  -0.822447   0.200334  -4.105
    InstTypeUni      -0.435290   0.126598  -3.438
    logAmount         0.536188   0.084074   6.378
                     Pr(>|z|)
    Genderf          0.025027 *
    DivisionDiv 2    0.000438 ***
    DivisionDiv 3    0.962709
    PercentFemale0.5 0.387759
    PercentFemale1   0.003862 **
    Age              0.137757
    IsContinuation1  2.34e-10 ***
    InstTypeOther    0.007233 **
    InstTypeUAS/UTE  4.04e-05 ***
    InstTypeUni      0.000585 ***
    logAmount        1.80e-10 ***
    ---
    Signif. codes:
    0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2    3.094      1.124   2.752
    2|3    5.342      1.112   4.805
    3|4    6.901      1.114   6.193
    4|5    8.379      1.120   7.478
    5|6   10.693      1.135   9.422
```

Odds Ratios and Confidence intervals:

Table 11: OR and corresponding Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.78	0.63	0.97
DivisionDiv 2	1.60	1.23	2.08
DivisionDiv 3	1.01	0.79	1.29
PercentFemale0.5	0.72	0.34	1.51
PercentFemale1	1.41	1.12	1.78
Age	1.01	1.00	1.02
IsContinuation1	2.09	1.66	2.62
InstTypeOther	0.54	0.35	0.85
InstTypeUAS/UTE	0.44	0.30	0.65
InstTypeUni	0.65	0.50	0.83
logAmount	1.71	1.45	2.02

Ordinal Regression for the Applicant Track

Summary of the final model:

```
formula:
ApplicantTrack ~ Gender + Division + PercentFemale + IsContinuation + InstType + Semester + logAmou
data:    internal_regression_data

 link  threshold nobs logLik   AIC     niter
 logit flexible  1623 -2015.18 4074.36 6(0)
 max.grad cond.H
 8.98e-07 5.3e+05

Coefficients:
                               Estimate
Genderf                        -0.11293
DivisionDiv 2                   0.31714
DivisionDiv 3                  -0.42406
PercentFemale0.5               -0.38864
PercentFemale1                  0.74102
IsContinuation1                 0.45664
InstTypeOther                  -0.81089
InstTypeUAS/UTE                -1.36995
InstTypeUni                    -0.38640
SemesterOct                    -0.15198
logAmount                       0.77310
Genderf:DivisionDiv 2          -0.61647
Genderf:DivisionDiv 3          -0.48649
DivisionDiv 2:PercentFemale0.5 -0.47640
DivisionDiv 3:PercentFemale0.5 -0.14851
DivisionDiv 2:PercentFemale1   -0.72419
DivisionDiv 3:PercentFemale1   -0.26792
                               Std. Error z value
Genderf                           0.17342  -0.651
DivisionDiv 2                     0.15845   2.002
DivisionDiv 3                     0.15984  -2.653
PercentFemale0.5                  0.53455  -0.727
```

```
PercentFemale1                   0.17834   4.155
IsContinuation1                  0.11667   3.914
InstTypeOther                    0.23809  -3.406
InstTypeUAS/UTE                  0.20413  -6.711
InstTypeUni                      0.12925  -2.990
SemesterOct                      0.09799  -1.551
logAmount                        0.09005   8.585
Genderf:DivisionDiv 2            0.29392  -2.097
Genderf:DivisionDiv 3            0.26215  -1.856
DivisionDiv 2:PercentFemale0.5   0.98345  -0.484
DivisionDiv 3:PercentFemale0.5   0.90013  -0.165
DivisionDiv 2:PercentFemale1     0.29309  -2.471
DivisionDiv 3:PercentFemale1     0.29780  -0.900
                               Pr(>|z|)
Genderf                         0.51491
DivisionDiv 2                   0.04533 *
DivisionDiv 3                   0.00798 **
PercentFemale0.5                0.46721
PercentFemale1                 3.25e-05 ***
IsContinuation1                9.08e-05 ***
InstTypeOther                   0.00066 ***
InstTypeUAS/UTE                1.93e-11 ***
InstTypeUni                     0.00279 **
SemesterOct                     0.12092
logAmount                       < 2e-16 ***
Genderf:DivisionDiv 2           0.03596 *
Genderf:DivisionDiv 3           0.06349 .
DivisionDiv 2:PercentFemale0.5  0.62809
DivisionDiv 3:PercentFemale0.5  0.86896
DivisionDiv 2:PercentFemale1    0.01348 *
DivisionDiv 3:PercentFemale1    0.36830
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2    3.146      1.274   2.470
2|3    5.775      1.147   5.035
3|4    7.539      1.140   6.615
4|5    9.699      1.147   8.458
5|6   11.903      1.161  10.250
```

Odds Ratios and Confidence intervals:

Table 12: OR and corresponding Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.89	0.64	1.25
DivisionDiv 2	1.37	1.01	1.87
DivisionDiv 3	0.65	0.48	0.89
PercentFemale0.5	0.68	0.24	1.94
PercentFemale1	2.10	1.48	2.98
IsContinuation1	1.58	1.26	1.99
InstTypeOther	0.44	0.28	0.71
InstTypeUAS/UTE	0.25	0.17	0.38
InstTypeUni	0.68	0.53	0.88
SemesterOct	0.86	0.71	1.04
logAmount	2.17	1.82	2.59
Genderf:DivisionDiv 2	0.54	0.30	0.96
Genderf:DivisionDiv 3	0.61	0.37	1.03
DivisionDiv 2:PercentFemale0.5	0.62	0.09	4.35
DivisionDiv 3:PercentFemale0.5	0.86	0.15	5.09
DivisionDiv 2:PercentFemale1	0.48	0.27	0.86
DivisionDiv 3:PercentFemale1	0.76	0.43	1.37

Ordinal Regression for Ranking

Summary of the final model:

```
formula:
Ranking ~ Gender + PercentFemale + Division + IsContinuation + PreviousRequest + InstType + logAmour
data:    internal_regression_data

 link  threshold nobs logLik   AIC     niter
 logit flexible  1623 -2305.97 4643.94 7(0)
 max.grad cond.H
 7.37e-12 5.1e+05

Coefficients:
                 Estimate Std. Error z value
Genderf          -0.30985    0.11009  -2.814
PercentFemale0.5 -0.29477    0.36906  -0.799
PercentFemale1    0.27939    0.11925   2.343
DivisionDiv 2     0.23153    0.13271   1.745
DivisionDiv 3    -0.22028    0.12586  -1.750
IsContinuation1   0.84141    0.11480   7.329
PreviousRequest1  0.20979    0.13204   1.589
InstTypeOther    -0.73369    0.22738  -3.227
InstTypeUAS/UTE  -0.99497    0.20053  -4.962
InstTypeUni      -0.43536    0.12620  -3.450
logAmount         0.56949    0.08423   6.761
                 Pr(>|z|)
Genderf          0.004887 **
PercentFemale0.5 0.424460
PercentFemale1   0.019132 *
```

```
DivisionDiv 2    0.081057 .
DivisionDiv 3    0.080084 .
IsContinuation1  2.32e-13 ***
PreviousRequest1 0.112090
InstTypeOther    0.001252 **
InstTypeUAS/UTE  6.99e-07 ***
InstTypeUni      0.000561 ***
logAmount        1.37e-11 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2    2.511      1.112   2.258
2|3    5.266      1.086   4.850
3|4    6.882      1.088   6.324
4|5    8.451      1.095   7.717
5|6   10.677      1.110   9.622
```

Odds Ratios and Confidence intervals:

Table 13: OR and Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.73	0.59	0.91
PercentFemale0.5	0.74	0.36	1.53
PercentFemale1	1.32	1.05	1.67
DivisionDiv 2	1.26	0.97	1.64
DivisionDiv 3	0.80	0.63	1.03
IsContinuation1	2.32	1.85	2.91
PreviousRequest1	1.23	0.95	1.60
InstTypeOther	0.48	0.31	0.75
InstTypeUAS/UTE	0.37	0.25	0.55
InstTypeUni	0.65	0.51	0.83
logAmount	1.77	1.50	2.09

A.5 Relative importance of each criteria

Relative Importance of the Different Steps

```
Call:
glm(formula = board_data1$IsApproved ~ Age + IsContinuation +
    Ranking + OverallGrade, family = "binomial", data = board_data1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7320  -0.3191   0.1762   0.5300   2.8416

Coefficients:
                 Estimate Std. Error z value
(Intercept)      2.313293   0.562387   4.113
Age             -0.027227   0.010607  -2.567
IsContinuation1  0.417989   0.202879   2.060
Ranking.L        4.296499   0.519097   8.277
Ranking.Q       -1.825581   0.406553  -4.490
Ranking.C       -0.139144   0.267250  -0.521
OverallGrade.L   1.417271   0.387013   3.662
OverallGrade.Q   0.032103   0.295225   0.109
OverallGrade.C   0.003249   0.170272   0.019
                Pr(>|z|)
(Intercept)     3.90e-05 ***
Age              0.01026 *
IsContinuation1  0.03937 *
Ranking.L        < 2e-16 ***
Ranking.Q       7.11e-06 ***
Ranking.C        0.60261
OverallGrade.L   0.00025 ***
OverallGrade.Q   0.91341
OverallGrade.C   0.98478
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2243.91  on 1622  degrees of freedom
Residual deviance:  976.92  on 1614  degrees of freedom
AIC: 994.92

Number of Fisher Scoring iterations: 6
```

Relative importance within the External step

Summary of the final model:

```
formula:
OverallGrade ~ Gender + Division + PercentFemale + ProposalCombined + ApplicantTrack + IsContinuatio
data:    ex_data

 link  threshold nobs logLik  AIC     niter
 logit flexible  1623 -835.55 1715.10 8(0)
 max.grad cond.H
 6.06e-08 8.6e+04

Coefficients:
                   Estimate Std. Error z value
Genderf           -0.033217   0.164301  -0.202
DivisionDiv 2     -0.006084   0.176129  -0.035
DivisionDiv 3     -0.265977   0.175834  -1.513
PercentFemale     -0.252904   0.280882  -0.900
ProposalCombined2 -0.621470   2.606549  -0.238
ProposalCombined3  3.246491   2.613563   1.242
ProposalCombined4  6.902134   2.625957   2.628
ProposalCombined5 10.883992   2.632709   4.134
ProposalCombined6 14.241967   2.647700   5.379
ApplicantTrack2   -0.735909   3.601075  -0.204
ApplicantTrack3    1.363007   3.573974   0.381
ApplicantTrack4    3.027672   3.577682   0.846
ApplicantTrack5    4.682749   3.582632   1.307
ApplicantTrack6    6.494427   3.586928   1.811
IsContinuation1    0.378674   0.166547   2.274
PreviousRequest1   0.092218   0.189470   0.487
SemesterOct        0.053376   0.132832   0.402
                  Pr(>|z|)
Genderf            0.83978
DivisionDiv 2      0.97245
DivisionDiv 3      0.13037
PercentFemale      0.36791
ProposalCombined2  0.81155
ProposalCombined3  0.21417
ProposalCombined4  0.00858 **
ProposalCombined5 3.56e-05 ***
ProposalCombined6 7.49e-08 ***
ApplicantTrack2    0.83807
ApplicantTrack3    0.70293
ApplicantTrack4    0.39740
ApplicantTrack5    0.19119
ApplicantTrack6    0.07021 .
IsContinuation1    0.02299 *
PreviousRequest1   0.62646
SemesterOct        0.68781
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Threshold coefficients:
    Estimate Std. Error z value
1|2   -3.895      4.395  -0.886
2|3    2.333      4.440   0.525
3|4    7.304      4.459   1.638
4|5   13.199      4.468   2.954
5|6   18.696      4.476   4.177
```

Odds Ratios and Confidence intervals:

Table 14: OR and corresponding Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.97	0.70	1.33
DivisionDiv 2	0.99	0.70	1.40
DivisionDiv 3	0.77	0.54	1.08
PercentFemale	0.78	0.45	1.35
ProposalCombined2	0.54	0.01	56.08
ProposalCombined3	25.70	0.42	2936.13
ProposalCombined4	994.39	15.68	116684.79
ProposalCombined5	53316.02	824.47	6346815.41
ProposalCombined6	1531820.52	22697.09	188330819.34
ApplicantTrack2	0.48	0.00	135.49
ApplicantTrack3	3.91	0.02	1203.10
ApplicantTrack4	20.65	0.13	6412.90
ApplicantTrack5	108.07	0.65	33976.46
ApplicantTrack6	661.45	3.94	210272.82
IsContinuation1	1.46	1.05	2.03
PreviousRequest1	1.10	0.76	1.59
SemesterOct	1.05	0.81	1.37

Relative importance of criteria within the Internal step

Summary of the final model:

```
formula:
Ranking ~ Gender + Division + PercentFemale + ProjectAssessment + ApplicantTrack + IsContinuation +
data:    data

 link  threshold nobs logLik  AIC     niter
 logit flexible  1623 -903.45 1852.89 8(0)
 max.grad cond.H
 9.63e-10 1.5e+04

Coefficients:
                   Estimate Std. Error z value
Genderf            -0.13388    0.15488  -0.864
DivisionDiv 2      -0.38768    0.16597  -2.336
DivisionDiv 3      -0.32121    0.17131  -1.875
PercentFemale0.5   -0.09689    0.54083  -0.179
PercentFemale1     -0.25787    0.16968  -1.520
ProjectAssessment2  2.89349    0.58765   4.924
ProjectAssessment3  5.72942    0.61107   9.376
ProjectAssessment4 10.34641    0.65075  15.899
ProjectAssessment5 14.50192    0.68990  21.020
ProjectAssessment6 19.73523    0.84756  23.285
ApplicantTrack2     2.53814    1.67610   1.514
ApplicantTrack3     2.37557    1.64346   1.445
ApplicantTrack4     3.15947    1.64503   1.921
ApplicantTrack5     4.15085    1.64861   2.518
ApplicantTrack6     5.55011    1.66405   3.335
IsContinuation1     0.55162    0.17049   3.235
PreviousRequest1    0.28234    0.17894   1.578
SemesterOct        -0.22630    0.13094  -1.728
                   Pr(>|z|)
Genderf            0.387379
DivisionDiv 2      0.019497 *
DivisionDiv 3      0.060784 .
PercentFemale0.5   0.857818
PercentFemale1     0.128572
ProjectAssessment2 8.49e-07 ***
ProjectAssessment3  < 2e-16 ***
ProjectAssessment4  < 2e-16 ***
ProjectAssessment5  < 2e-16 ***
ProjectAssessment6  < 2e-16 ***
ApplicantTrack2    0.129946
ApplicantTrack3    0.148326
ApplicantTrack4    0.054781 .
ApplicantTrack5    0.011810 *
ApplicantTrack6    0.000852 ***
IsContinuation1    0.001214 **
PreviousRequest1   0.114596
SemesterOct        0.083948 .
---
Signif. codes:
```

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2    2.008      1.640   1.225
2|3    6.697      1.739   3.852
3|4   11.229      1.754   6.403
4|5   16.688      1.775   9.404
5|6   22.650      1.814  12.483
```

Odds Ratios and Confidence intervals:

Table 15: OR and corresponding Confidence Intervals

	OR	2.5 %	97.5 %
Genderf	0.87	0.65	1.190000e+00
DivisionDiv 2	0.68	0.49	9.400000e-01
DivisionDiv 3	0.73	0.52	1.010000e+00
PercentFemale0.5	0.91	0.32	2.580000e+00
PercentFemale1	0.77	0.55	1.080000e+00
ProjectAssessment2	18.06	5.94	6.086000e+01
ProjectAssessment3	307.79	96.85	1.084170e+03
ProjectAssessment4	31145.10	9046.79	1.179575e+05
ProjectAssessment5	1986560.46	533373.72	8.088255e+06
ProjectAssessment6	372306626.97	74022174.88	2.073676e+09
ApplicantTrack2	12.66	0.59	4.766000e+02
ApplicantTrack3	10.76	0.54	3.868000e+02
ApplicantTrack4	23.56	1.17	8.472200e+02
ApplicantTrack5	63.49	3.14	2.297250e+03
ApplicantTrack6	257.27	12.30	9.525120e+03
IsContinuation1	1.74	1.24	2.430000e+00
PreviousRequest1	1.33	0.93	1.880000e+00
SemesterOct	0.80	0.62	1.030000e+00