Replication Package

Grantmaking, Grading on a Curve, and the Paradox of Relative Evaluation in Nonmarkets

Jérôme Adda and Marco Ottaviani

The analysis in this paper is based on data from the European Research Council (ERC) that reports all funded applications across all disciplines for the years 2007-2020. A second dataset consists of funding applications to the Research Council Norway (RCN) for the years 2002-2019.

The appendix below contains a description of the files with the name of the variables and summary statistics. In line with the user agreement, the original data files from the RCN must be eliminated from personal computers after the completion of the project. Researchers interested in the full replication of the research should contact the RCN directly and apply for the micro data. We provide a thorough description of the source files and the variables in the appendix. We provide intermediary files allowing the replication of all figures and tables in the paper.

Sources of data:

ERC dashboard webpage:
https://dashboard.tech.ec.europa.eu/qs_digit_dashboard_mt/public/sense/app/afe00964-3272-45c4-b60c-b64ed20d98d1/sheet/61a0bd1d-cd6d-4ac8-8b55-80d8661e44c0/state/analysis

RCN contact:
post@forskningsradet.no
https://www.forskningsradet.no/en/about-the-research-council/Contact/

Softwares:
We used Stata 17 and Python to produce the results.

Files in replication package

File name	Software	Description	Origin
Code files			
Run_35905.do	Stata	Main file to produce the tables and figures	Authors
Shape_35905.do	Stata	Shapes the raw RCN data files	Authors
r_i_text.R	Python	Predict the ERC panel based on application detail	Authors
preprocess_bert.R	Python	Predict the ERC panel based on application detail	Authors
merge_pred.R	Python	Predict the ERC panel based on application detail	Authors
final_bert_class.py	Python	Predict the ERC panel based on application detail	Authors
Data files			
Success rates .xlsx	Excel	Success rate at ERC	ERC dashboard webpage
erc2020_complete.xlsx	Excel	Funded projects at ERC	ERC dashboard webpage
reliability.dta	Stata	inter-rater agreement measures by ERC panel, derived from RCN data	Constructed by Shape_35905.do
RCN_Predicted.dta	Stata	Predicted probabilities of classification of applications into ERC panels	Constructed by Shape_35905.do

Replication:

Run the Stata file *Run_35905.do.*

This file produces the Table I and the Figures I, VIII and IX in the paper, as well as Tables II, IV and V and Figures XIV and XV in the online appendix. The figures are in pdf format, the tables in Tex.

The file *Run_35905.do* takes as input the files listed below:

- Success rates .xlsx
- erc2020_complete.xlsx
- reliability.dta
- RCN_Predicted.dta

All these files are in the replication package.

Note: Table V in the appendix requires the raw data to be computed. Due to the confidentiality of the RCN data, this table is produced by the file Shape_35905.do and requires the access to the raw RCN data.

Shaping the raw RCN files

The file Shape_35905.do shapes the raw files from the RCN to construct reliability.dta and RCN_Predicted.dta, that serve as input to the main file *Run_35905.do.*
It also produces the first panel of Table II in the appendix and the bootstrap results in Table V in the appendix.

It draws on numerous files that are provided by the RCN and listed below. Those interested in the full replication should enquire at the RCN to gain access to them. The appendix of this note lists all the content of these files for easier retrieving.

Additional data files used by Shape_35905.do

File name	Description
soknader2020fagkoder.xls	RCN applications for 2020
OUT-5433-v6.xls	Application grades 2020-2021
out-5433-2012-2011.xls	Application grades 2011-2012
0824_Søknader_2002-2019_til_DESREV-prosjektet_Hovedkarakter.xlsx	RCN grades 2002-2019
Boccini University - Researcher project + Young research talents.xls Boccini University - Other support FRIPRO + SFF.xls	RCN grades 2013-2019
0128_Desrev_søknadsfrist_2020_Søkere.xlsx	Abstract of application, 2020
0128_Desrev_søknadsfrist_2021_Søkere.xlsx	Abstract of application, 2021
0624_Søknader_2002-2006_til_DESREV-prosjektet.xlsx	RCN applicants
0624_Søknader_2007-2009_til_DESREV-prosjektet.xlsx	RCN applicants
0624_Søknader_2010-2012_til_DESREV-prosjektet.xlsx	RCN applicants
0624_Søknader_2013-2014_til_DESREV-prosjektet.xlsx	RCN applicants
0624_Søknader_2015-2016_til_DESREV-prosjektet.xlsx	RCN applicants
0624_Søknader_2017-2018_til_DESREV-prosjektet.xlsx	RCN applicants
0810_Søknader_2002-2019_til_DESREV-prosjektet_PL_kjønn.xlsx	Sex of RCN applicants
1209_Søknader_2002-2006_til_DESREV-prosjektet_PL_fødselsdato.xlsx	Date of birth of applicants
1209_Søknader_2007-2010_til_DESREV-prosjektet_PL_fødselsdato.xlsx	Date of birth of applicants
1209_Søknader_2011-2014_til_DESREV-prosjektet_PL_fødselsdato.xlsx	Date of birth of applicants
1209_Søknader_2015-2017_til_DESREV-prosjektet_PL_fødselsdato.xlsx	Date of birth of applicants
1209_Søknader_2018-2019_til_DESREV-prosjektet_PL_fødselsdato (1).xlsx	Date of birth of applicants
0624_Søknader_2002-2006_til_DESREV-prosjektet.xlsx	Referee information
0624_Søknader_2007-2009_til_DESREV-prosjektet.xlsx	Referee information
0624_Søknader_2010-2012_til_DESREV-prosjektet.xlsx	Referee information
0624_Søknader_2013-2014_til_DESREV-prosjektet.xlsx	Referee information
0624_Søknader_2015-2016_til_DESREV-prosjektet.xlsx	Referee information
0624_Søknader_2017-2018_til_DESREV-prosjektet.xlsx	Referee information
0624_Søknader_2019_til_DESREV-prosjektet.xlsx	Referee information

Assigning RCN applications to ERC panels

Each RCN application is assigned to a predicted ERC panel by the files:

- r_i_text.R
- preprocess_bert.R
- merge_pred.R
- final_bert_class.py

1. analysisRCN.csv and analysisRCN2020_2021New.csv are data files generated by Shape_35905.do.

There is two types of datasets:
(1) title + abstract already classified into the corresponding panel
(2) title + abstract with no classification
D(1) serve as input for our training. The model generated from the training is used to classify the entries of D(2).

2. We pre-process the texts by doing some simple tasks such as punctuation to make it suitable for the BERT algorithm. This is done on R by the code "preprocess_bert.R". It uses the csv datasets in the raw_data folder to generate 6 datasets that we save in the "input_folder".
"XX_training_data_app" and "XX_prediction_data_app.csv" where XX=LS,PE or SH.
"XX_training_data_app" is the training dataset that will be used by BERT to build the model. The model will be used to predict the entries of "XX_prediction_data_app.csv".

3. The BERT is executed by the Python script "final_bert_class.py". For each "XX", the result of the prediction will be saved in int_output_data as "XX_predictions.csv".

4. The last step is simply to attach to each prediction the project id and the text it corresponds to. This is done by the code "merge_pred.R" which saves the final version of the prediction in the folder "final_output_data" in the dataset "XX_predictions_ext.csv".

APPENDIX

Content of file OUT-5433-v6.xls

Description:

Observations: 30,651

Variables: 39

Variable name	Storage type	Variable label
aktivitet	str15	aktivitet
søknadsfristår	int	søknadsfrist år
prosjektnummer	long	prosjektnummer
søknadstype	str15	søknadstype
søknadstypevariant	str34	søknadstypevariant
beslutning	str8	beslutning
søkt	int	søkt
bevilget	int	bevilget
tema	byte	tema
fagomrade	str99	fagomrade
healthcategory	str34	health category
healthresearchactivity	str24	health research activity
prosjektlederId	long	prosjektlederId
prosjektlederFodselsår	int	prosjektlederFodselsår
prosjektlederKjønn	str6	prosjektlederKjønn
prosjektlederhjemland	str7	prosjektleder hjemland
prosjektlederarbeidsland	str7	prosjektleder arbeidsland
prosjektlederland	str2	prosjektleder land
ekspertFodselsår	int	ekspertFodselsår
ekspertKjønn	str6	ekspertKjønn
eksperthjemland	str3	ekspert hjemland
ekspertarbeidsland	str7	ekspert arbeidsland
ekspertland	str2	ekspert land
ekspertId	str7	ekspertId
vurderingstype	str17	vurderingstype
FormidlingogkommunikasjonFP	byte	Formidling og kommunikasjon | FP
ForskningskvalitetFP	str1	Forskningskvalitet | FP
GjennomføringFP	str1	Gjennomføring | FP
Gjennomføringsplanogressursbeh	byte	Gjennomføringsplan og ressursbehov | FP
ProsjektgruppenUF	byte	Prosjektgruppen | UF
ProsjektkvalitetforFoUprosjek	byte	Prosjektkvalitet for FoU-prosjektet | IPN
ProsjektlederUF	byte	Prosjektleder | UF
Prosjektlederogprosjektgruppen	byte	Prosjektleder og prosjektgruppen | FP
Relevansiforholdtilutlysning	byte	Relevans i forhold til utlysningen | FP
Samletvurderingfrafageksperte	str1	Samlet vurdering fra fageksperten/panelet
VirkningerogeffekterFP	str1	Virkninger og effekter | FP
VirkningerogeffekterFPMobi	str1	Virkninger og effekter | FP-Mobilitet
Vitenskapeligdristighetogfagl	byte	Vitenskapelig dristighet og faglig fornyelse
VitenskapeligkvalitetFP	byte	Vitenskapelig kvalitet | FP

Content of file 0824_Søknader_2002-2019_til_DESREV-prosjektet_Hovedkarakter.xlsx

Description:

Observations: 98,462

Variables: 2

Variable name	Storage type	Variable label
Prosjektnummer	long	Prosjektnummer
Hovedkarakter	byte	Hovedkarakter

Content of files

- **soknader2020fagkoder.xls**
- **soknader2021fagkoder.xls**

Description:

Observations: 2,363

Variables: 37

Variable name	Storage type	Variable label
Prosjnr	long	Prosj.nr.
Prosjtittel	str144	Prosj.tittel
Prosjansv	str143	Prosj.ansv.
Prosjleder	str46	Prosj.leder
ProgAkt	str9	Prog/Akt
Utlysning	str52	Utlysning
Søknadstype	str15	Søknadstype
Variant	str32	Variant
Tema1	str96	Tema 1
Tema2	str96	Tema 2
Tema3	str96	Tema 3
Tema4	str58	Tema 4
Tema5	str37	Tema 5
Fagkode1	str65	Fagkode 1
Fagkode2	str64	Fagkode 2
Fagkode3	str64	Fagkode 3
Fagkode4	str64	Fagkode 4
Fagkode5	str64	Fagkode 5
Fradato	str10	Fra dato
Tildato	str10	Til dato
Søkt1år	int	Søkt 1. år
Søkt2år	int	Søkt 2. år
Søkt3år	int	Søkt 3. år
Søkt4år	int	Søkt 4. år
Søkt5år	int	Søkt 5. år
Søkttotalt	int	Søkt totalt
Hovedkarakter	byte	Hovedkarakter
Behstatus	str21	Beh.status
Beslutning	str8	Beslutning
Saksb	str5	Saksb.
M1	str4	M1
M2	str48	M2
M3	str2	M3
M4	str59	M4
M5	str5	M5
M6	str65	M6
Kommentar	str1839	Kommentar

Content of files

- **0128_Desrev_søknadsfrist_2020_Søkere.xlsx**
- **0128_Desrev_søknadsfrist_2021_Søkere.xlsx**

Description:

Observations: 4,407

Variables: 11

Variable name	Storage type	Variable label
Prosjektnummer	long	Prosjektnummer
OrganisasjonsId	long	Organisasjons-Id
Prosjektansvarlignavnsiste	str86	Prosjektansvarlig navn siste
Kortnavn	str14	Kortnavn
Organisasjonsnummer	str9	Organisasjonsnummer
FoUsektor	str15	FoU sektor
FoUkategori	str44	FoU kategori
Prosjektledernavnsiste	str39	Prosjektleder navn siste
PersonId	long	Person-Id
Epost	str42	Epost
Prosjektsammendrag	str2038	Prosjektsammendrag

Content of file

- **Boccini University - Researcher project + Young research talents.xls**

Description:

Observations: 64,789

Variables: 38

Variable name	Storage type	Variable label
name	str14	aktivitet
aktivitet	int	søknadsfrist år
søknadsfristår	long	prosjektnummer
prosjektnummer	str20	søknadstype
søknadstype	str9	søknadstypevariant
søknadstypevariant	str13	beslutning
beslutning	long	søkt
søkt	long	bevilget
bevilget	str69	tema
tema	str99	fagomrade
fagomrade	str87	health category
healthcategory	str81	health research activity
healthresearchactivity	long	prosjektlederId
prosjektlederId	int	prosjektlederFodselsår
prosjektlederFodselsår	str6	prosjektlederKjønn
prosjektlederKjønn	str7	prosjektleder hjemland
prosjektlederhjemland	str7	prosjektleder arbeidsland
prosjektlederarbeidsland	str2	prosjektleder land
prosjektlederland	int	ekspertFodselsår
ekspertFodselsår	str6	ekspertKjønn
ekspertKjønn	str7	ekspert hjemland
eksperthjemland	str7	ekspert arbeidsland
ekspertarbeidsland	str2	ekspert land
ekspertland	str7	ekspertId
ekspertId	str17	vurderingstype
vurderingstype	str1	Formidling og kommunikasjon | FP
FormidlingogkommunikasjonFP	str1	Forskningskvalitet | FP
ForskningskvalitetFP	str1	Gjennomføring | FP
GjennomføringFP	str1	Gjennomføringsplan og ressursbehov | FP
Gjennomføringsplanogressursbeh	str1	Prosjektgruppen | UF
ProsjektgruppenUF	str1	Prosjektkvalitet for FoU-prosjektet | IPN
ProsjektkvalitetforFoUprosjek	str1	Prosjektleder | UF
ProsjektlederUF	str1	Prosjektleder og prosjektgruppen | FP
Prosjektlederogprosjektgruppen	str1	Relevans i forhold til utlysningen | FP
Relevansiforholdtilutlysning	str1	Samlet vurdering fra fageksperten/panelet
Samletvurderingfrafageksperte	str1	Virkninger og effekter | FP
VirkningerogeffekterFP	str1	Vitenskapelig dristighet og faglig fornyelse
Vitenskapeligdristighetogfagl	str1	Vitenskapelig kvalitet | FP

Content of file

- **0624_Søknader_2002-2006_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2007-2009_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2010-2012_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2013-2014_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2015-2016_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2017-2018_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2019_til_DESREV-prosjektet.xlsx**

Description:

Observations: 28,683; 15,747; 16,913; 11,727; 12,633; 6,303

Variables: 18

Variable name	Storage type	Variable label
Prosjektnummer	long	Prosjektnummer
Hovedaktivitet	str44	Hovedaktivitet
Aktivitetskodeognavn	str91	Aktivitetskode og navn
Prosjekttype	str19	Prosjekttype
Søknadstype	str41	Søknadstype
Søknadstypevariant	byte	Søknadstypevariant
Prosjekttittelprimærspråk	str148	Prosjekttittel primærspråk
Prosjektperiodensfradato	int	Prosjektperiodens fra-dato
Prosjektperiodenstildato	long	Prosjektperiodens til-dato
Søknadsdato	double	Søknadsdato
Søknadsår	int	Søknadsår
Prosjektansvarlignavnsiste	str79	Prosjektansvarlig navn siste
Prosjektledernavnsiste	str40	Prosjektleder navn siste
PersonId	long	Person-Id
Prosjektsammendrag	strL	Prosjektsammendrag
Prosjektfase	str10	Prosjektfase
Søktbeløp	double	Søkt beløp
Revidertbudsjett	double	Revidert budsjett

Content of file 0810_Søknader_2002-2019_til_DESREV-prosjektet_PL_kjønn.xlsx

Description:

Observations: 98,463

Variables: 3

Variable name	Storage type	Variable label
Prosjektnummer	long	Prosjektnummer
PersonId	long	Person-Id
Kjønn	str6	Kjønn

Content of files

- **1209_Søknader_2002-2006_til_DESREV-prosjektet_PL_fødselsdato.xlsx**
- **1209_Søknader_2007-2010_til_DESREV-prosjektet_PL_fødselsdato.xlsx**
- **1209_Søknader_2011-2014_til_DESREV-prosjektet_PL_fødselsdato.xlsx**
- **1209_Søknader_2015-2017_til_DESREV-prosjektet_PL_fødselsdato.xlsx**
- **1209_Søknader_2018-2019_til_DESREV-prosjektet_PL_fødselsdato (1).xlsx**

Description:

Observations: 27,741; 20,629; 22,568; 19,251; 12,428

Variables: 9

Variable name	Storage type	Variable label
Prosjektnummer	type	Prosjektnummer
Prosjektetsfraår	long	Prosjektets fra-år
Prosjektetstilår	int	Prosjektets til-år
Rolle	int	Rolle
PersonId	str13	Person-Id
Fødselsdato	long	Fødselsdato
Kjønn	double	Kjønn
Fraår	str6	Fra år
Tilår	int	Til år

Content of files

- **0624_Søknader_2002-2006_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2007-2009_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2010-2012_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2013-2014_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2015-2016_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2017-2018_til_DESREV-prosjektet.xlsx**
- **0624_Søknader_2019_til_DESREV-prosjektet.xlsx**

Description:

Observations: 33,422; 24,963; 45,042; 34,578; 40,479; 43,280; 22,656

Variables: 8

Variable name	Storage type	Variable label
Prosjektnummer	long	Prosjektnummer
Rolle	str10	Rolle
Rollenavn	str55	Rollenavn
Kjønn	str6	Kjønn
PersonId	long	Person-Id
Fødselsdato	double	Fødselsdato
Tittel	str32	Tittel
Akademiskgrad	str10	Akademiskgrad