

**The intersecting forces of prescriptivism and diachronic shifts:
A multivariate study of four years of university admissions essays**

Lars Hinrichs, The University of Texas at Austin

“Probabilistic variation across dialects and varieties” workshop Leuven,
05 April 2016

Prescriptivists are most successful in influencing language use when they prescribe what is happening anyway (Hinrichs et al. 2015):

There is a statistical drift in language use away from bookish/formal and toward colloquial/informal options (Mair & Leech 2006).

Therefore, when you tell people to “use **that** (not **which**) as restrictive relativizer”; or to “use the active voice (not the passive)”, then you are **prescribing the informal choice**. And you are likely to be successful because you are aligning with a drift that is already underway.



from ***Le Petit Prince*** by Antoine de Saint-Exupéry

In short: writers treat precepts distinctly - each variable has its own story and statistical trajectory.

An interesting dataset:

University admissions essays

from "a large public university in the Southern U.S."

Data: admissions essays

- 28,958,705 words
- 51,825 essays
- 25,975 writers (~2/student)
- admitting years 2004-2007

- includes only essays from successful applications
- many students “auto-admitted”
- every student writes on 2 topics chosen from a range options:

greatest challenge

ideal classmate

architecture

sci method

misc

fiction

- annotated for the student's background, bio
- numeric measures of academic success:
ACT, SAT, GPA for each college year
- each essay rated by a human reader

What can we hope to study in the data?

- correctness ideologies held by students
(& parents, ghostwriters...)
- strategies to impress

>> two indexical orders of prescriptive language-use

Related research

- linguistic prescriptivism and student essays
- linguistic consequences of the pressure to impress as a psychological state
- familiarity with (near-)automated scoring
(Perelman 2014)

Perelman:

Students know that SAT essay scores favor

- long sentences
- high word count
- concise statements/
absence of ambiguity
- words like ***indubitably***
- high lexical density

EDUCATION | ON EDUCATION

Facing a Robo-Grader? Just Keep Obfuscating Mellifluously

By MICHAEL WINERIP APRIL 22, 2012

Research questions

- Which prescriptivism-related variables predict academic success?
- Which of the Perelman-strategies correlate with essay scores, academic success?*

*Pennebaker et al. (2014) show that **CDI** — an aggregate measure of POS frequencies — predicts academic success (i.e. mean-GPA) about as well as SAT score.

Method

- POS frequencies and word frequencies
(Tauszcik & Pennebaker 2010)
- Use of features from the prescriptive canon
(Peters & Young 1997; Hinrichs et al. 2015)

>> descriptive & inferential statistics;
aggregate data analysis

Data was annotated **at the essay-level** for 280 variables and/or aggregated to **means per writer**

- 90 vectors of information about student background, biography (from registrar)
- 100 vectors of POS-frequencies and semantic field representation (from Pennebaker)
- we added 90 vectors with interest in prescriptivism in mind (our addition)

Ethical considerations:

The data, already anonymized, was not to leave Pennebaker's hard drive.

Statistical challenge:
working with aggregated data
(Szmrecsanyi 2013)

ANALYSES

- 1.** Descriptive: individual variables
- 2.** Descriptive:
indices “defy”, “impress”, “comply”
- 3.** Predictive

Correlations

Indices

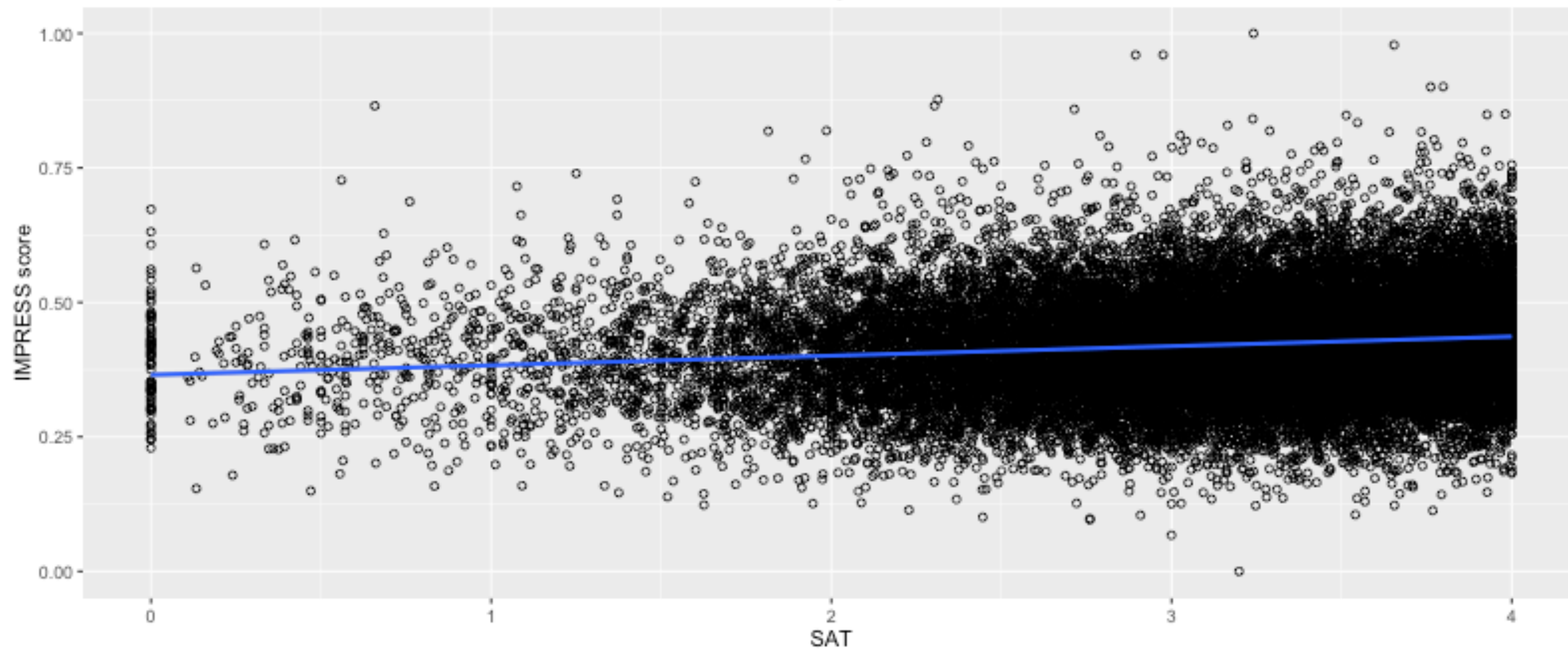
DEFY

defy =
splitInfinitive + contraction +
finalPreposition + restrictiveWhich
+ hopefully + passive

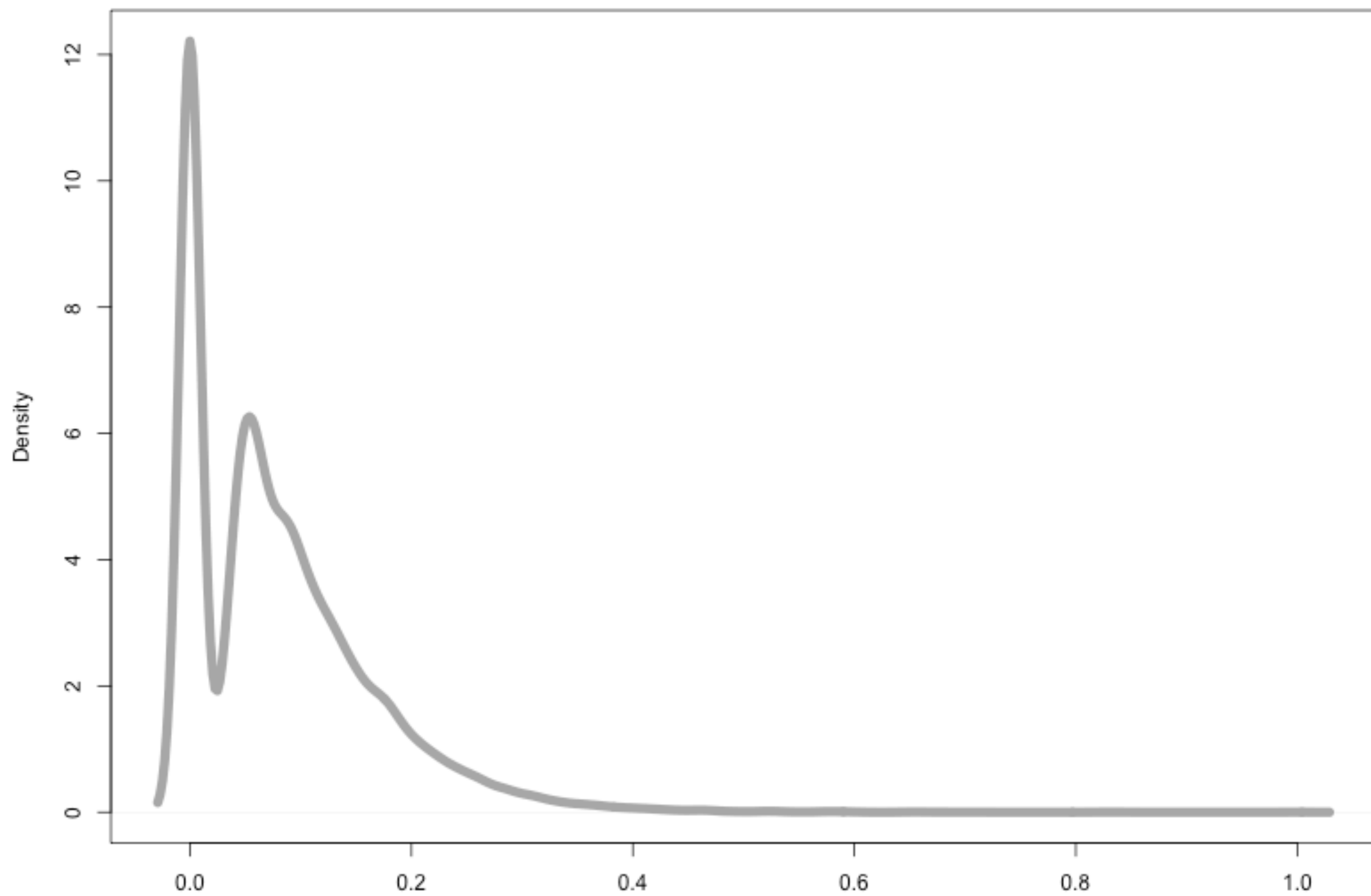
IMPRESS

impress =
rarewords + whom + TTR + nouns +
sentences

IMPRESS by SAT-total

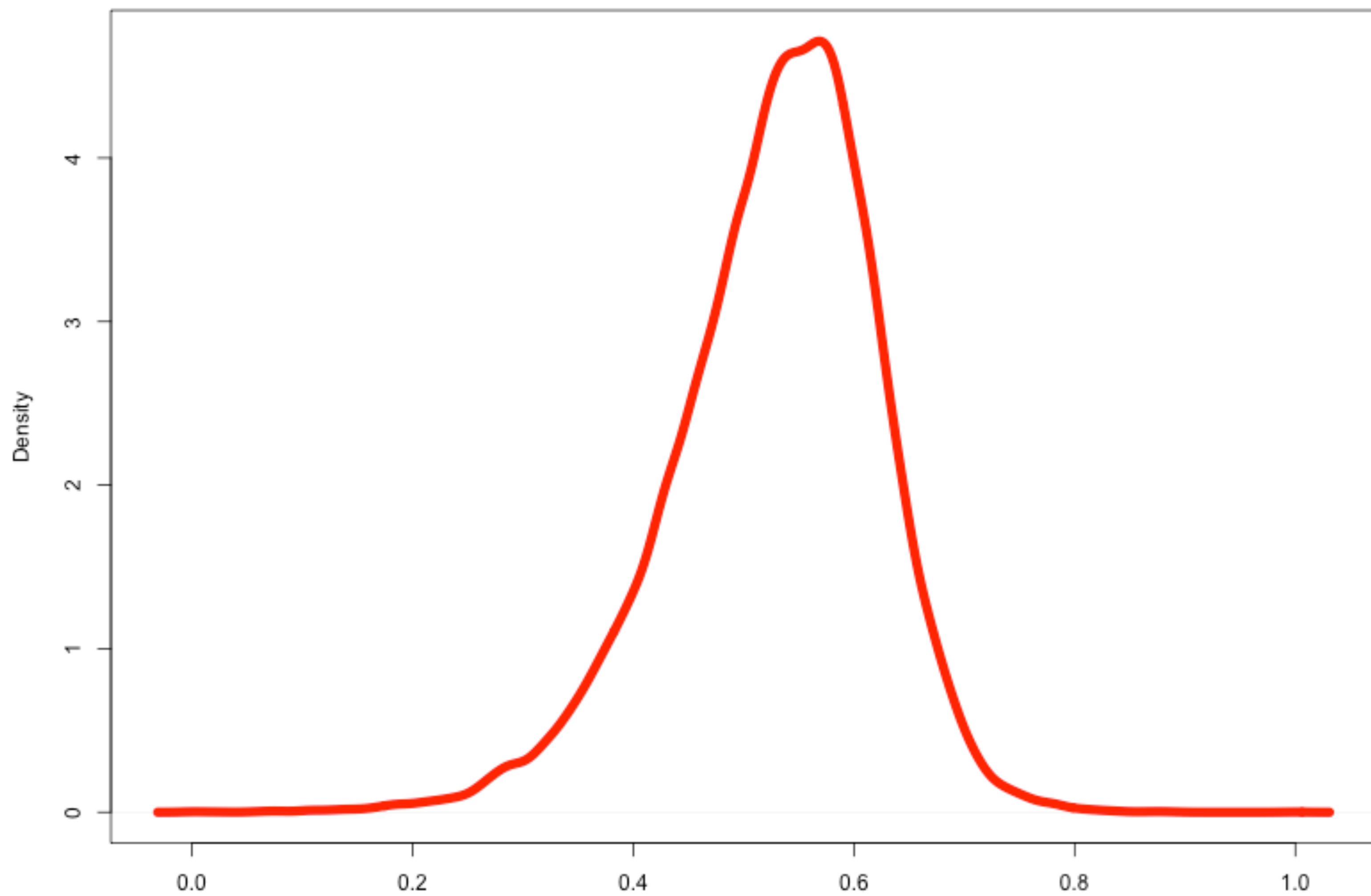


Density for COMPLY1 (positives only)



N = 25974 Bandwidth = 0.009693

Density for COMPLY2 (positives - negatives)

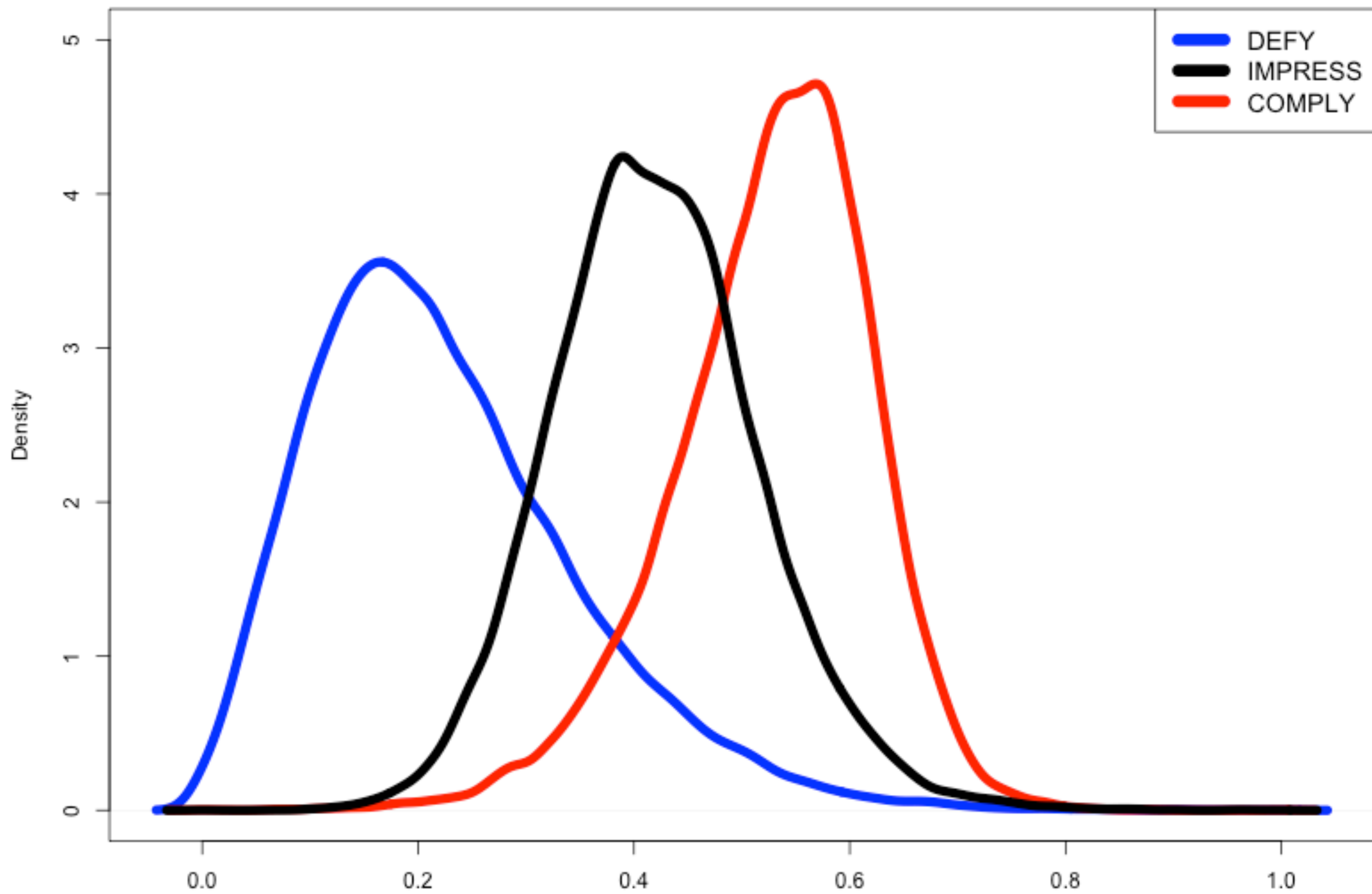


N = 25974 Bandwidth = 0.01032

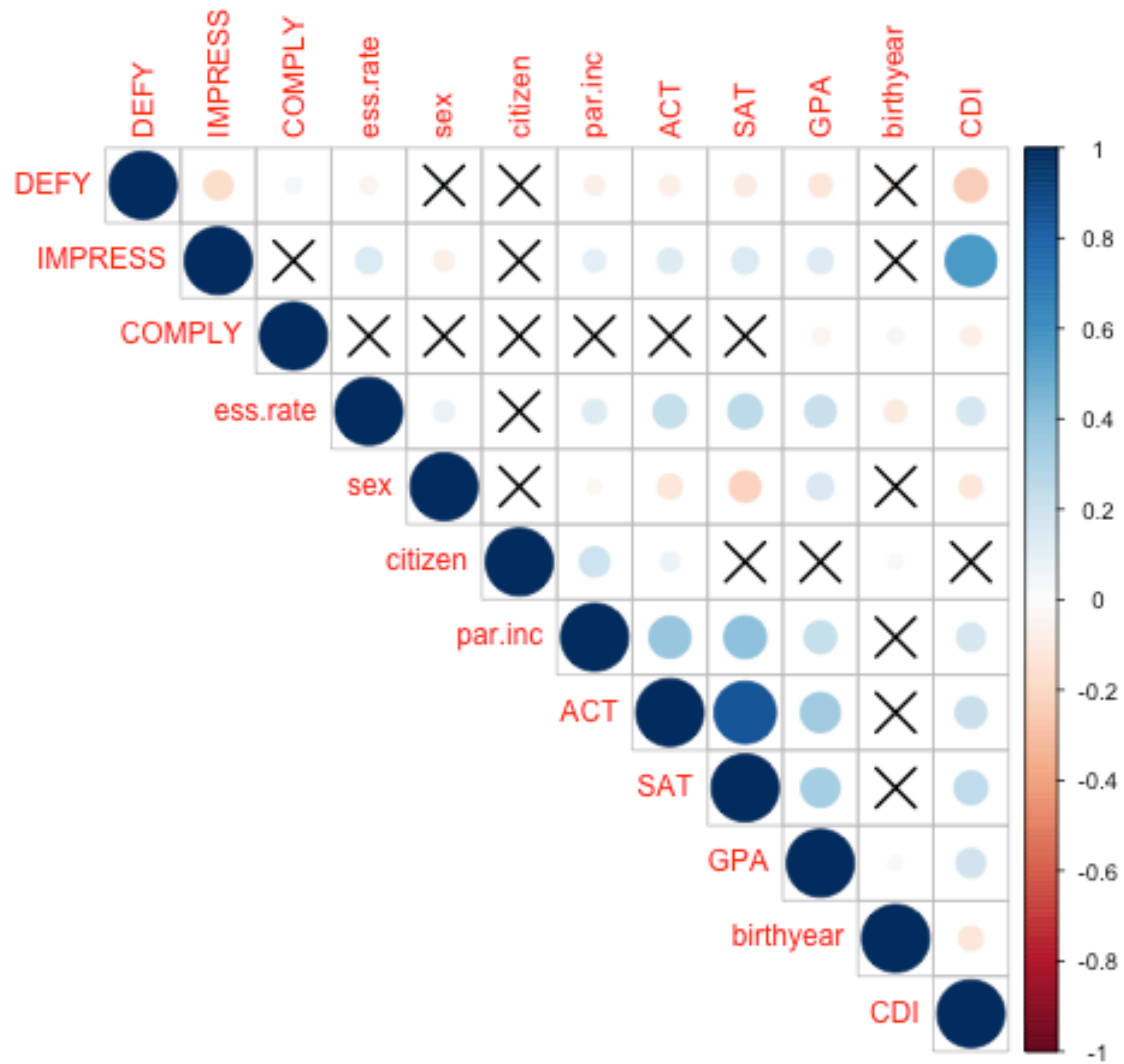
COMPLY

comply =
pipedPiping + restrictiveThat -
splitInfinitive - contraction -
finalPreposition - restrictiveWhich
- hopefully - passive

Density for Style Indices



N = 25974 Bandwidth = 0.01418



Hierarchical modeling

[1]	"writer_id"	"words"	"sentences"	"school"	"major"
[6]	"txhs"	"autoadmt"	"essay_1"	"essay_2"	"avgEssayRate"
[11]	"txres"	"citizen"	"sex"	"ethnic"	"parinc"
[16]	"actengl"	"actmath"	"actread"	"actsci"	"acttot"
[21]	"satverb"	"satmath"	"sattot"	"satequiv"	"hs_code"
[26]	"fathed"	"mothed"	"ori_name"	"meanGPA"	"GradIn4yrs"
[31]	"teaisd"	"tearate"	"teatype"	"ncescity"	"ncestit1"
[36]	"nceslat"	"nceslong"	"popzip"	"popurb"	"poprur"
[41]	"birthyear"	"birthyr"	"CDI"	"piedPiping"	"restrThat"
[46]	"rare"	"whom"	"TTR"	"nouns"	"splitInf"
[51]	"contractions"	"finalPrps"	"restrWhich"	"hopefully"	"comply"
[56]	"impress"	"passive"	"supercomply"	"defy"	

m1

outcome: COMPLY

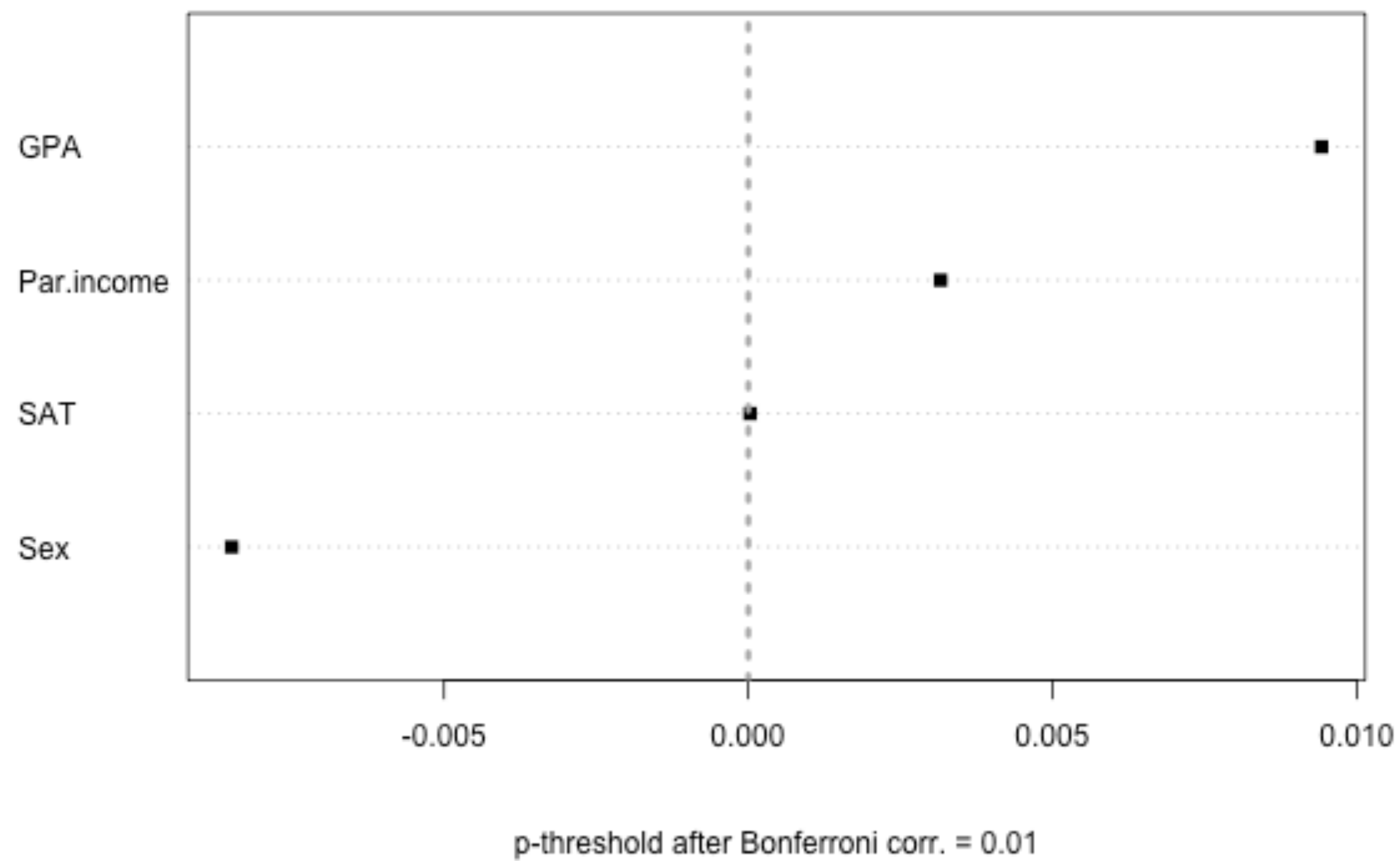
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: supercomply

	Chisq	Df	Pr(>Chisq)	
meanGPA	106.829	1	< 2.2e-16	***
sattot	45.878	1	1.258e-11	***
parinc	48.514	1	3.280e-12	***
sex	48.361	1	3.546e-12	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Significant predictors for COMPLY



m2

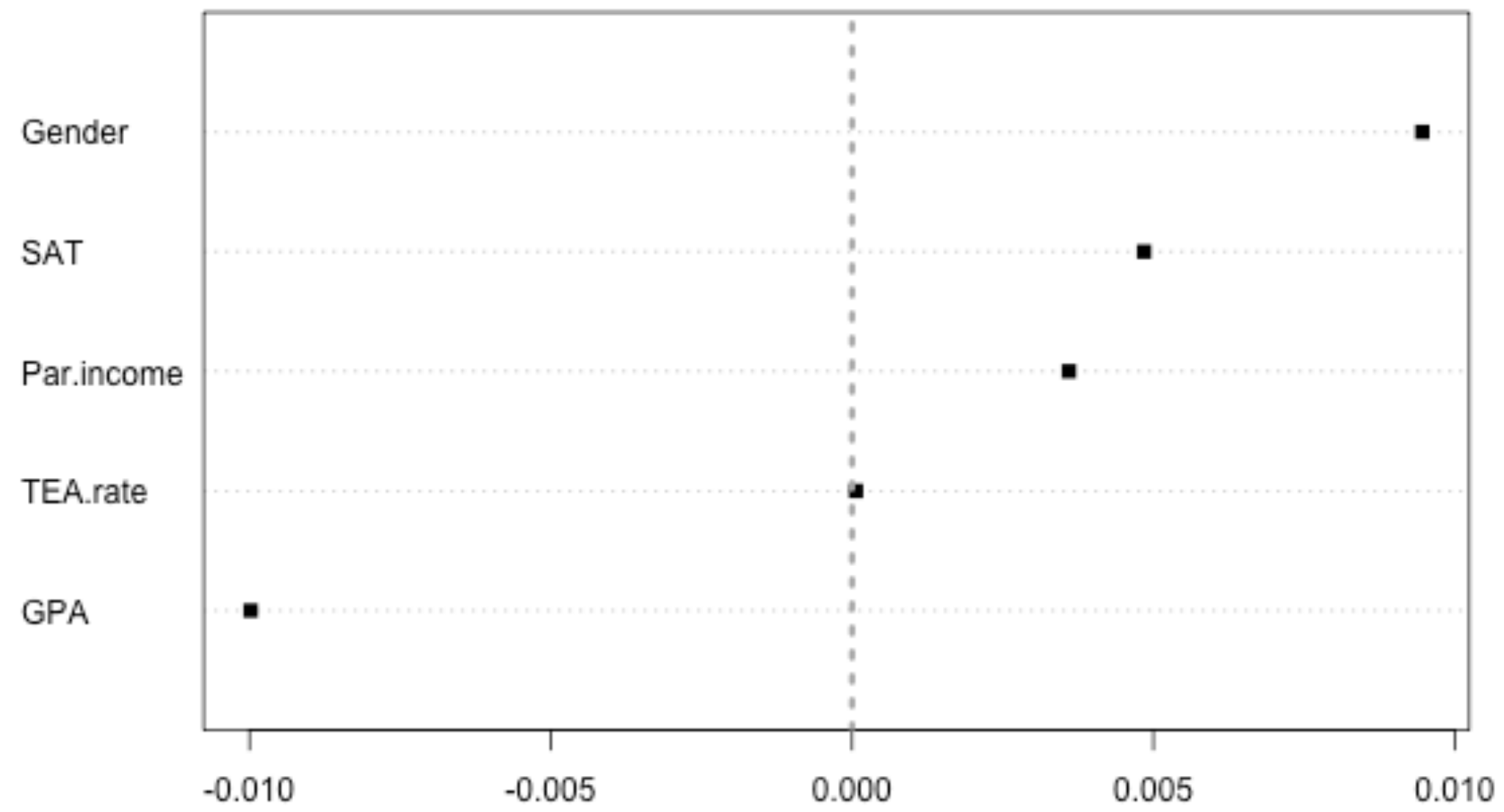
outcome: IMPRESS

Response: impress

	Chisq	Df	Pr(>Chisq)	
meanGPA	78.895	1	< 2.2e-16	***
sattot	152.864	1	< 2.2e-16	***
parinc	47.041	1	6.951e-12	***
sex	47.603	1	5.217e-12	***
tearate	18.047	1	2.156e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Significant predictors for IMPRESS



p-threshold after Bonferroni corr. = 0.01

CONCLUSIONS

- This statistical look at the admissions essay data has shown that the social indexicality of precept-**compliant** language use plays a significant part in the type of social success that is academic success.

- Recent discussions of prescriptivism have invested a lot of effort into the task of deconstructing the descriptivism-prescriptivism binary as a simple clash of good and bad.

Not only have many studies shown that prescriptivism can change language. The perspective taken in this paper also shows that the ability to deploy the indexical values of **compliant** language use is a core skill in the process of developing an academically/professionally/economically successful persona.

- The use of language that is designed to **impress** in many senses an even more socially effective skill.

- As linguists — whether socio- or other — we must be cognizant of the following facts about prescriptivism:
 - It changes language use.
 - Following its rules is a big piece of the puzzle that cements social class membership.
- Therefore, we ought not to dismiss prescriptivism.
- In terms of methodology, it seems that the ever-larger and often interesting datasets that we are increasingly likely to stumble upon are set to continue to challenge us to develop and refine our methods for aggregate data analysis. I have found correlations and meta-aggregation of variables into indices—in addition to methods already established in qt. linguistics such as regression—to be quite useful in this respect.

LH@utexas.edu
lars@texasenglish.org

Lars Hinrichs, The University of Texas at Austin
“Probabilistic variation across dialects and varieties” workshop
Leuven, 05 April, 2016

**The intersecting forces of prescriptivism and diachronic shifts:
A multivariate study of four years of university admissions essays**

Outline

- Introduction
 - The data
 - Background
 - Framework
 - Research questions
- Methods
 - POS frequencies and word frequencies
 - Features from the prescriptive canon
- Findings
- Discussion
 - Prescriptivism and equity in higher education: a critique
 - Toward a quantitative sociolinguistics of precept-uptake

References

- Grafmiller, Jason, Benedikt Szmrecsanyi & Lars Hinrichs. 2016. Restricting the restrictive relativizer: Constraints on subject and non-subject English relative clauses. *Corpus Linguistics and Linguistic Theory*. Under review.
- Hinrichs, Lars, Benedikt Szmrecsanyi & Axel Bohmann. 2015. Which-hunting and the Standard English relative clause. *Language* 91(4). 806–836.

- Lavergne, Gary M. & Bruce Walker. 2001. Developing a concordance between the ACT assessment and the SAT I: Reasoning test for the University of Texas at Austin.
- Mair, Christian & Geoffrey Leech. 2006. Current changes in English syntax. In B. Aarts & A. McMahon (eds.), *The handbook of English linguistics*, 318–342. Malden, MA: Blackwell.
- Pennebaker, James W. & Cindy K. Chung. 2005. Tracking the social dynamics of responses to terrorism: Language, behavior, and the Internet. In S. Wessely & V. Krasnov (eds.), *Psychological responses to the New Terrorism: A NATO-Russia dialogue*, 159–170. Amsterdam et al.: IOS.
- Pennebaker, James W., Cindy K. Chung, Joey Frazer, Gary M. Lavergne & David I. Beaver. 2014. When small words foretell academic success: The case of college admissions essays. *PloS one* 9(12). e115844. doi:10.1371/journal.pone.0115844.
- Perelman, Les. 2014. When “the state of the art” is counting words. *Assessing Writing* 21. 104–111.
- Peters, Pam & Wendy Young. 1997. English grammar and the lexicography of usage. *Journal of English Linguistics* 25(4). 315–331.
- Szmrecsanyi, Benedikt. 2013. Analyzing aggregated linguistic data. In M. Krug & J. Schlüter (eds.), *Research methods in language variation and change*, 433–455. Cambridge: CUP.
- Tausczik, Yla R. & James W. Pennebaker. 2010. The psychological meaning of Words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1). 24–54. doi:10.1177/0261927X09351676 (14 July, 2013).
- Wolk, Christoph & Benedikt Szmrecsanyi. 2016. Top-down and bottom-up advances in corpus-based dialectometry. In M. Côté, R. Knooihuizen & J. Nerbonne (eds.), *The future of dialects*, 225–244. Berlin: Language Science Press.

Contact:

LH@utexas.edu