Online Appendix

This online appendix accompanies the article "Why do Students Enroll in Political Science classes?" by Jonas B. Bunte published in *PS: Political Science & Politics.* Please note that references to Tables and Figures in the article are represented by Arabic numerals whereas references to Tables and Figures in this appendix are denoted by capitalized letters.

Contents

1 The Characteristics of the Sample

This section presents summary statistics describing the characteristics of the seven sections that were part of the experiment. The data show that the sections did not differ significantly across observable characteristics such as students' field of study, year of study, and gender. Recall that these are required classes that all students at this university must take in order to graduate. Taken together, the data indicate that class characteristics such as day of the week, time of day, and instructor identity are unlikely to have played a role.

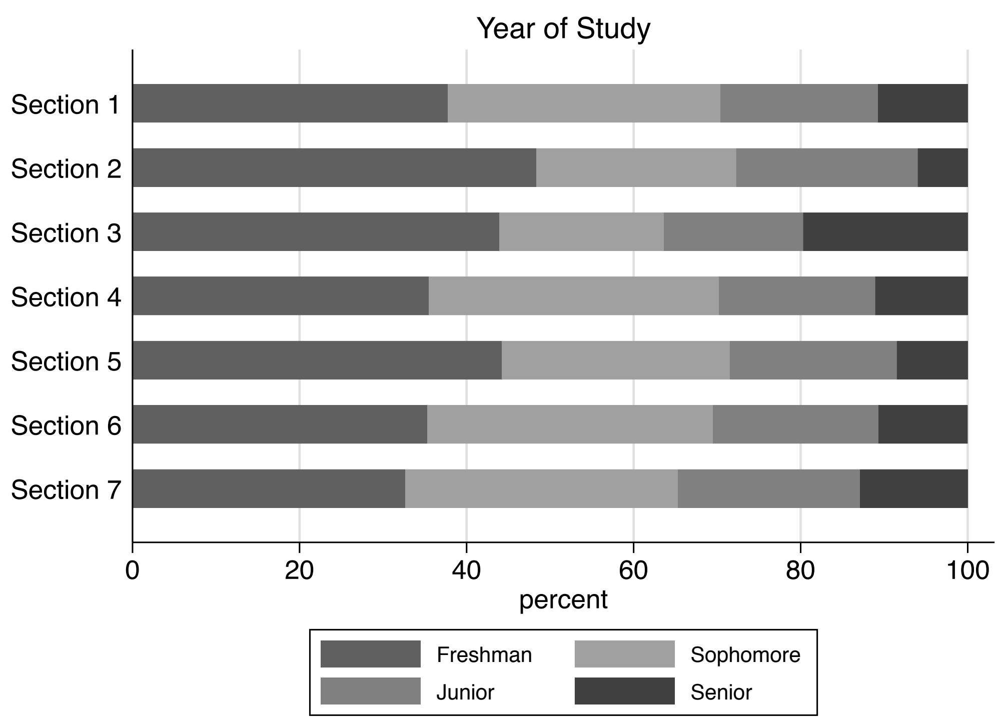


Figure A: Distribution of students' year of study across sections. The data indicate that the sections were similarly heterogeneous.

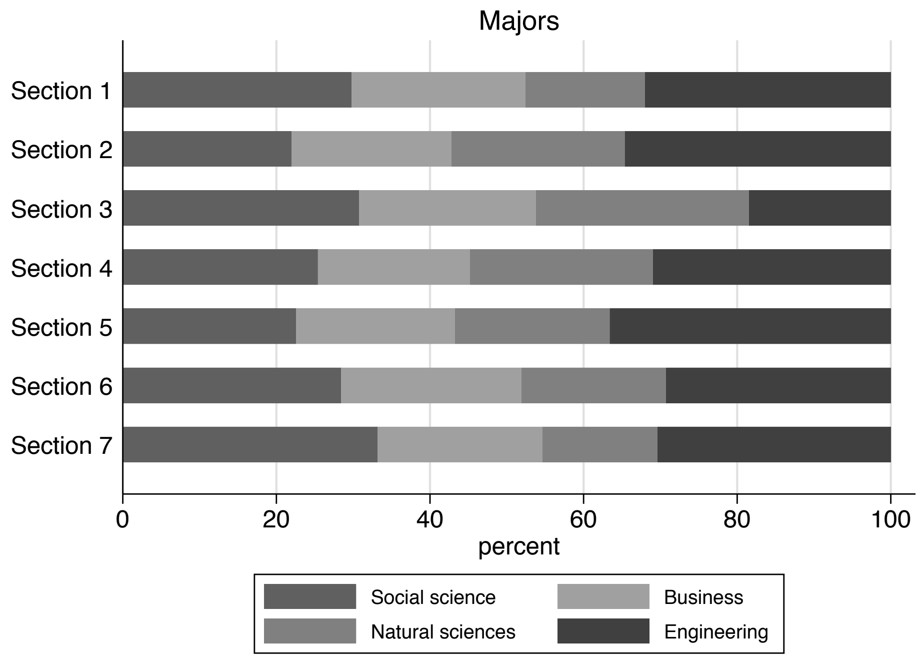


Figure B: Distribution of students' majors across sections. The data show that the distribution of majors did not differ significantly across sections.

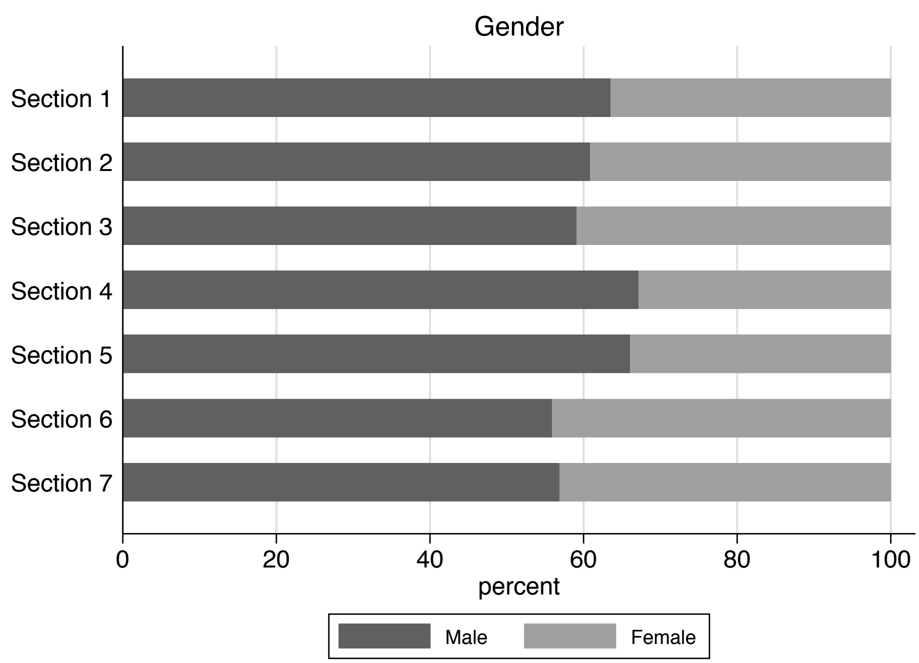


Figure C: Distribution of students' gender across sections. The data suggest that the gender distribution was very similar across sections.

2 Details regarding Treatment Design

2.1 Topics Treatment

The experiment involved advertising a class to be taught in the next semester. That class was a comparative politics class, effectively comparing policies in the United States, Germany, and Sweden. Due to the nature of the class, the topics treatment (see Figure 4) focused on comparative politics issues when advertising the future class. However, comparative politics is only one of several subfields in political science: it would also have been possible to expand the experiment to include classes from other subfields, such as political theory, international relations, or American politics. In fact, it might be argued that I should have designed several types of treatments, one for each subfield in Political Science, to examine which of these treatments is most successful at recruiting students. In my opinion, this is a question worthy of scholarly attention and should be subject to future research.

However, I did not pursue this line of thought in the current experiment as advertising multiple types of classes from different subfields presents two methodological challenges. First, it is possible that the logistics of presenting multiple topics (one per subfield) might result in bias. For example, when pitching multiple classes, I could advertise all classes myself; however, students might recognize that I am not an expert in, say, political theory. If they find my pitch for a political theory class less convincing than my pitch for a comparative politics class, resulting differences in enrollment might reflect this implicit bias rather than a 'true' effect. Alternatively, instead of advertising different classes myself, I could ask my colleagues who are experts in the respective fields to do the pitch themselves; however, in this case variation across effect sizes might be the result of differences in the personality, style, or prior popularity of each instructor rather than measuring differences across political science subfields. Either way, pitching several types of treatments drawn from different political science subfields would have likely resulted in biased estimates. For this reason, I focused on a research design that centered on a single instructor of a single class for which I am an expert.

Second, my empirical analysis uses matching algorithm to identify the average treatment effect. Increasing the number of treatments would also have necessitated a larger sample size. Increasing the sample size was not possible at my university, as there are only so many sections of this class offered in a given semester (and I visited them all). If I would have increased the number of treatments from 3 (skills, current events, and topics) to 6 (skills, current events, comparative topic, international relations topic, political theory topic, and American politics topic) the experiment would almost certainly have yielded insufficient observations for a matching analysis.

2.2 Skills Treatment

The field of political science is constantly evolving. One of the most prominent trends is the increasing use of statistics in political science. In light of this development, the treatment definition estimating the effect of skills could have emphasized quantitative literacy and data analytics skills development.

However, in designing this treatment, I decided against this for several reasons. First, I specifically did not include words such as "math" and "statistics" as I was afraid that they would scare students away, even if they were interested in learning skills useful for later careers. Instead, I used the wording of "conducting cost-benefit analyses" to capture the effect of students interested in a skill involving empirical analysis of quantitative data.

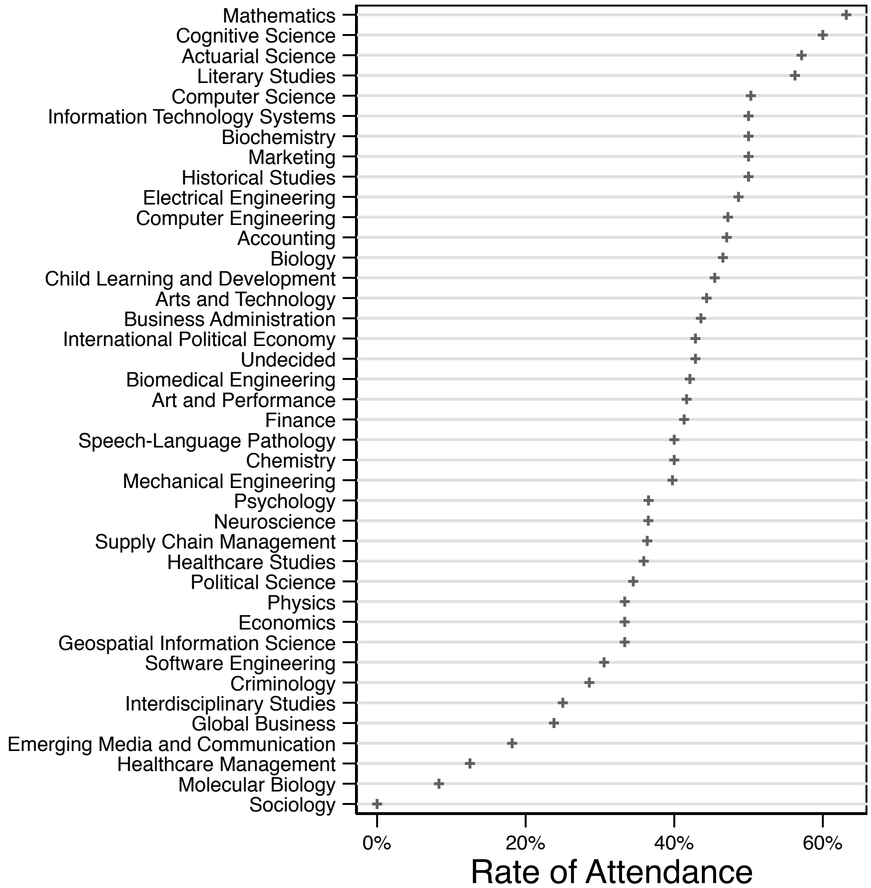
Second, I did not want to talk about "quantitative skills" in the abstract, hoping that students would themselves make the mental connections to realize how such skills would be applicable in their future careers. Instead, I wanted to ensure that my three-minute presentation would increase the likelihood of students immediately recognizing the value of the skills offered in the subsequent class. For this reason, I chose to present skills *in context* of specific fields of work.

3 Student Attendance and the Accuracy of Estimates

As noted in the manuscript, 840 of the 1,467 students enrolled in these sections did not attend class on the day I visited. It is possible that the patterns of class absences are not random and thus introduce bias. For instance, it is possible that students who are interested in politics might be more likely to attend class and thus may be more likely to enroll in subsequent political science classes. If this were the case, it would come as no surprise that the main reason as to why they would be interested in taking another political science class is to learn more about politics. My response to this potential threat to internal validity is twofold.

First, if this line of reasoning were accurate, we would expect the rate of absenteeism to be particularly low among political science majors: These are the students most likely interested in politics and therefore should be least likely to skip class. However, this does not appear to be the case: Figure D displays the percentage of students enrolled in a particular major that attended class on the day of the experiment. The data show that students majoring in Political Science are no less likely to skip class than other majors. In fact, only 34% of Political Science majors enrolled in the Government sections were attending class. In contrast, 63% of students studying Mathematics, 50% of Bioengineers, and 48% of Electrical Engineers attended class, even though their majors are less related to politics. These data suggest that students interested in politics (as proxied by their declared major) are *not* more likely to attend class than students presumably less interested in politics.

One reason might be that it is precisely the students already interested in politics that did not show up for class. After all, as these government classes offer an easy introduction to American government, students interested in politics are likely to already know the information presented in class. As a result — and contrary to the reasoning above — students interested in politics might be *less* likely to attend class. This would



Note: The graph displays the percentage of students enrolled in a particular major that attended class on the day of the experiment.

Figure D: Attendance by major

imply — if there is bias due to non-attendance — that the analyses underestimate the true effect sizes: The treatments might have been even more effective if the students receiving treatments would have been more interested in politics.

Second, if this line of argument were accurate, we would expect that among the students exposed to treatment, political science majors should be particularly likely to enroll in the advertised class in the next semester. However, this does not appear to be the case when analyzing the set of students exposed to treatment that subsequently enrolled in the class next semester. Of these students, only 27% had a declared major related to politics (either Political Science or International Political Economy). Instead, the majority of the students that enrolled in the subsequent class were majoring in fields unrelated to politics (Natural sciences 21%, computer engineering 35%, economics 13%, undecided 4%). These data suggest that, conditional on receiving treatment, those students with an prior interest in politics (as proxied by their major) were no more likely to enroll than other students.

4 Assessing the Performance of the Matching Algorithm

The findings presented in the article represent the average treatment effects after matching. That is, I use propensity score matching to identify student pairs that are as similar as possible across observable characteristics, but that received different treatments. I use three covariates for matching: a) students' field of study (social science, business and management, natural sciences, and engineering), b) students' year of study (freshmen, sophomores, juniors, and seniors), and c) students' gender. Matching ensure that I compare the observed outcome of a student in one treatment group with the outcome of a second student with identical observable characteristics who received a different treatment.

This section presents evidence suggesting that the matching approach performs extraordinarily well. The balance plots shown below summarize the distribution of treatment versus control groups across the three variables for each model. Their interpretation is straightforward: The closer solid and dotted lines, the more comparable are treatment and control groups. The evidence suggests that in every model, matching improved the balance of covariates.

4.1 Analysis 1

The first analysis compares students exposed to a presentation to a section of students that I did not visit. The balance plots below compare the treatment and the control groups for each treatment across the three covariates used for matching. The graphs illustrate that matching significantly improves the balance of covariates across treatment and control groups.

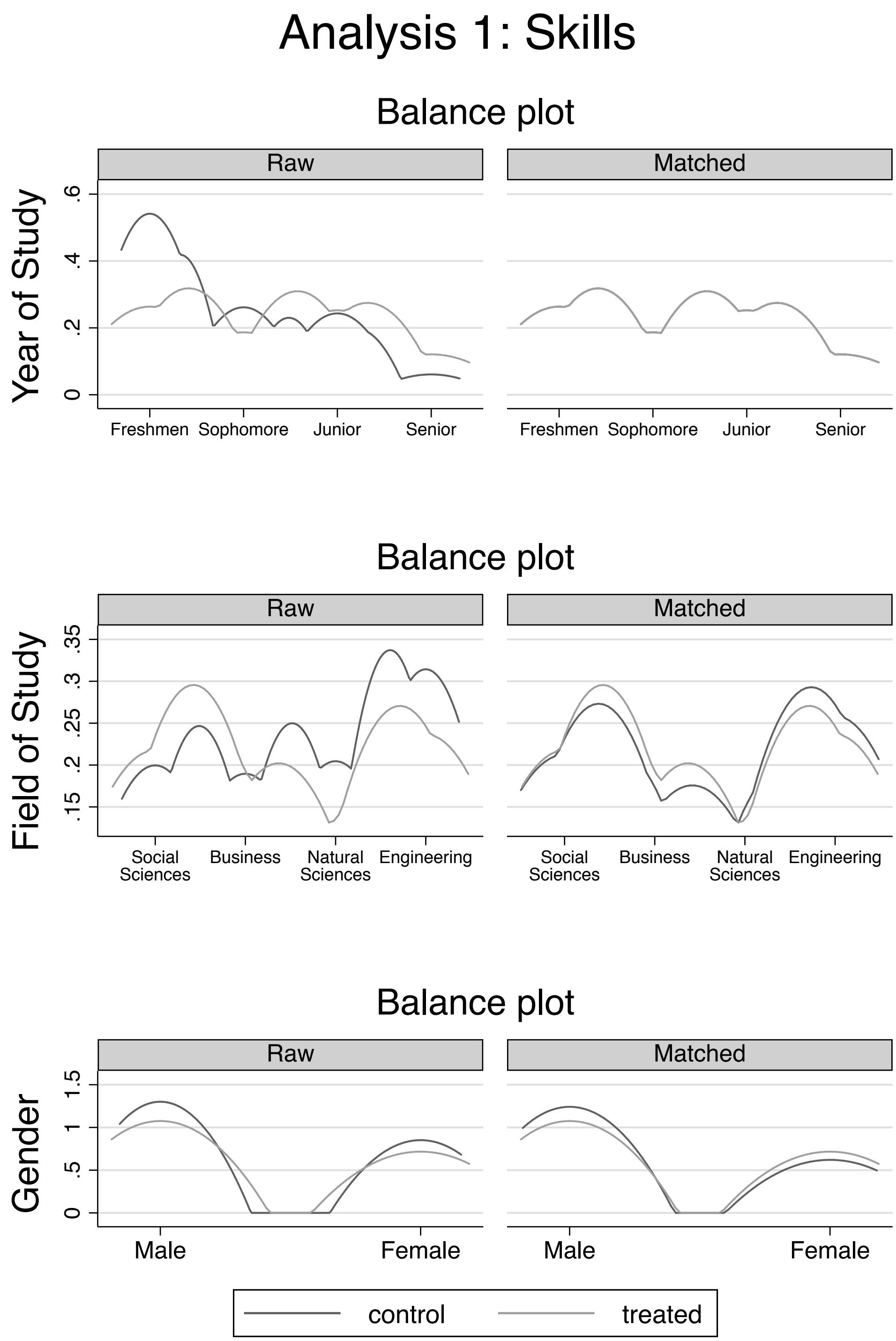


Figure E: Matching performance of skill model.

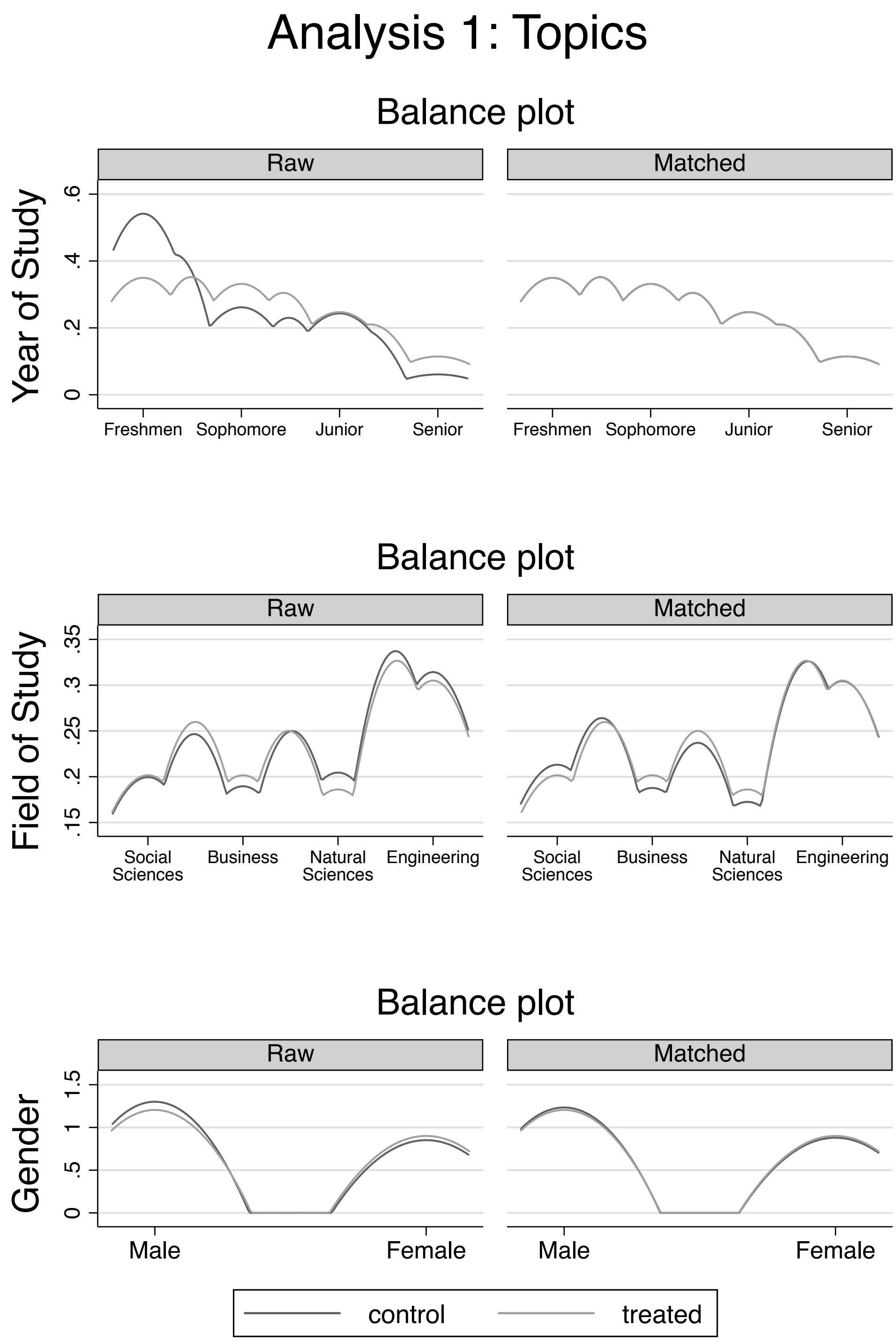


Figure F: Matching performance of topics model.

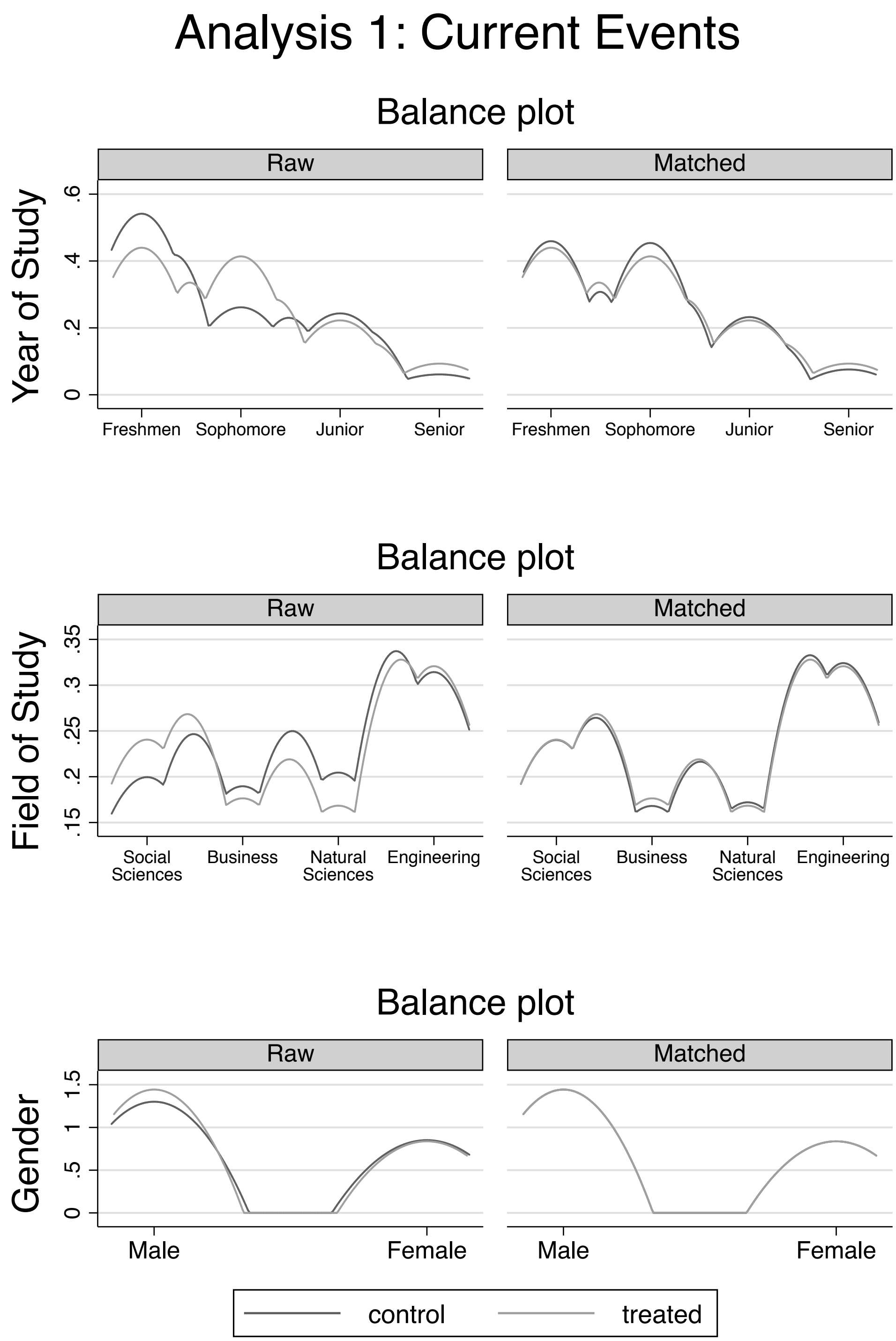


Figure G: Matching performance of current event model.

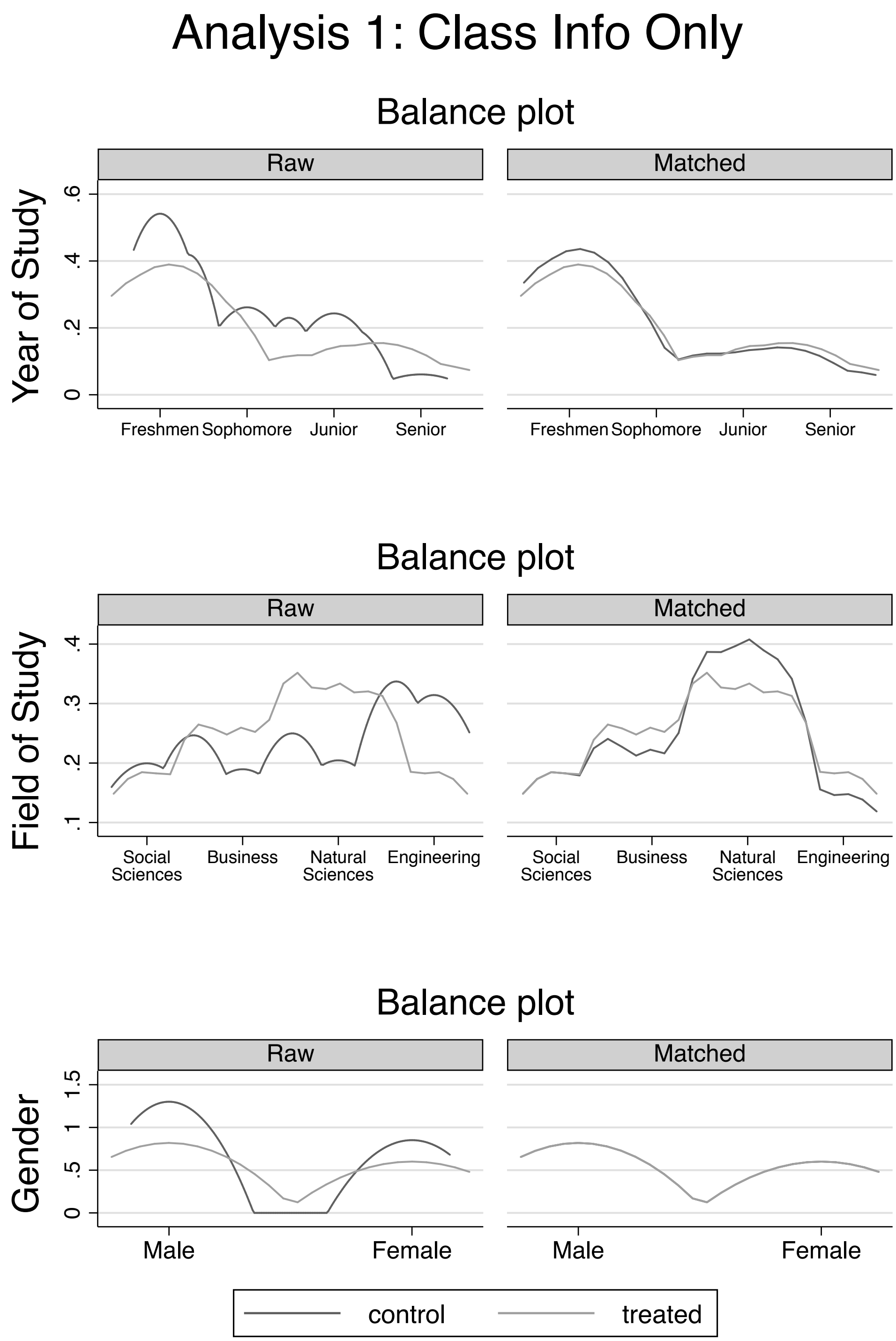


Figure H: Matching performance of class information model.

4.2 Analysis 2

The second analysis compares students that attended the class I visited to those that were absent that day. The balance plots below compare the treatment and the control groups for each treatment across the three covariates used for matching. The graphs illustrate that matching significantly improves the balance of covariates across treatment and control groups.

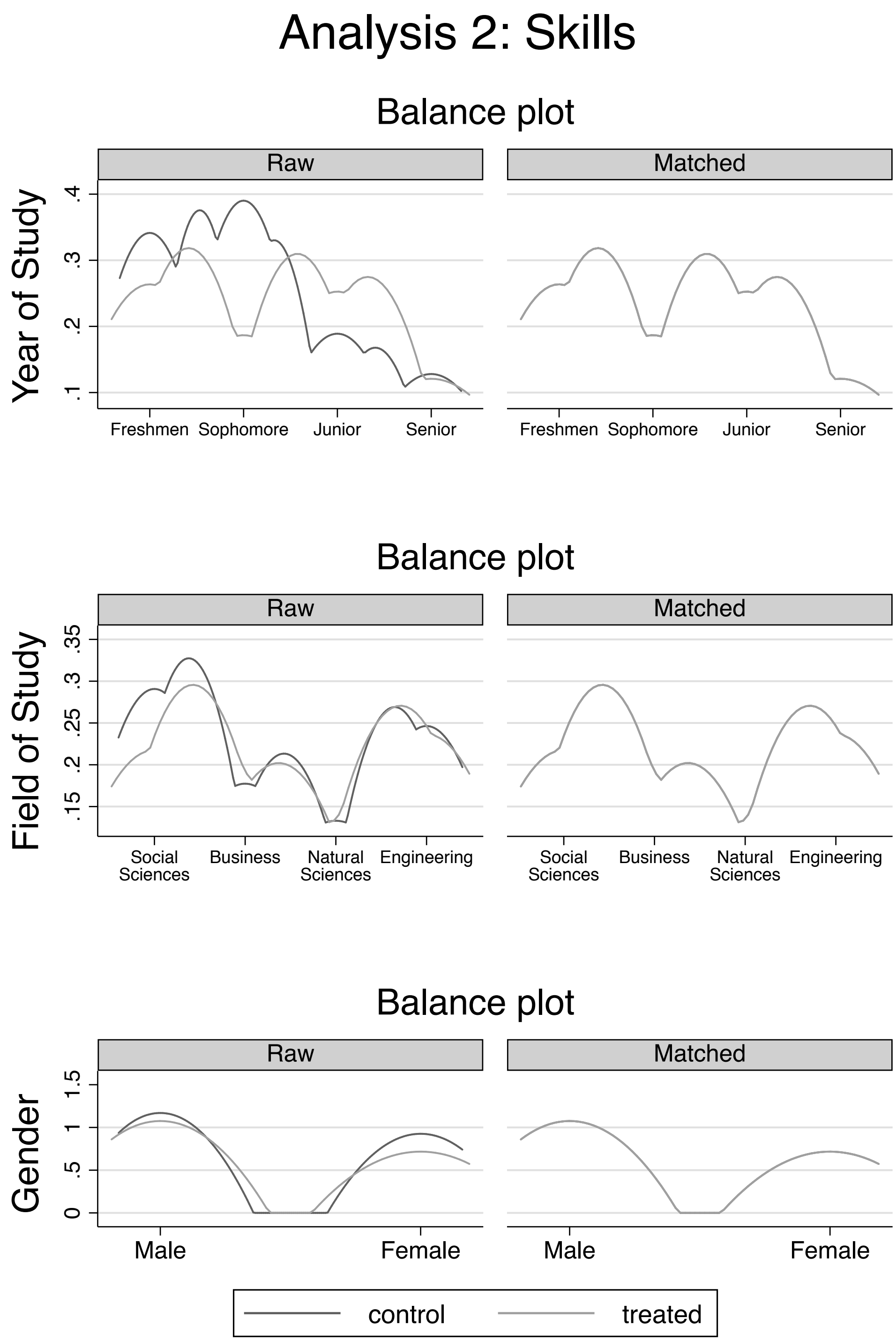


Figure I: Matching performance of skill model.

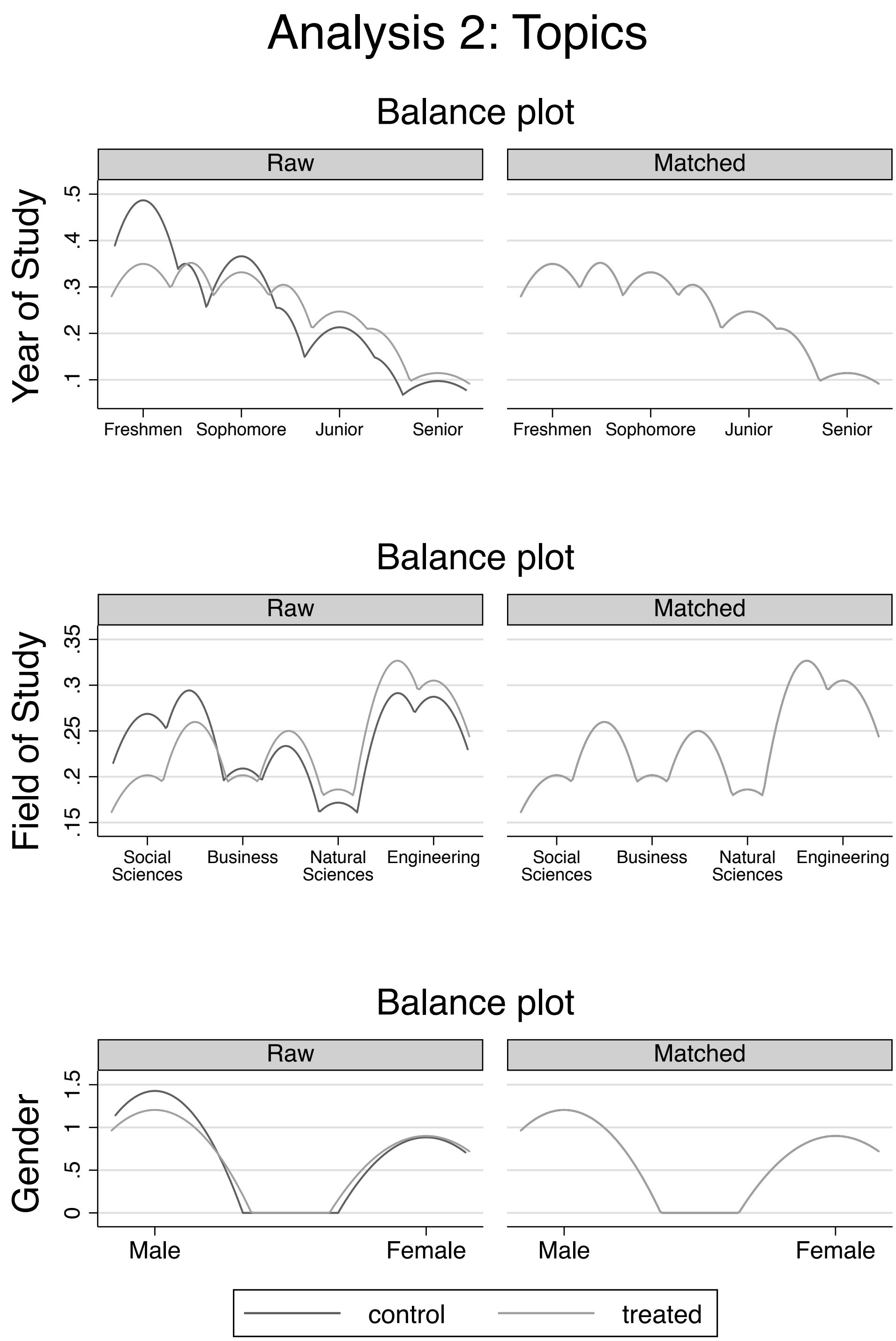


Figure J: Matching performance of topics model.

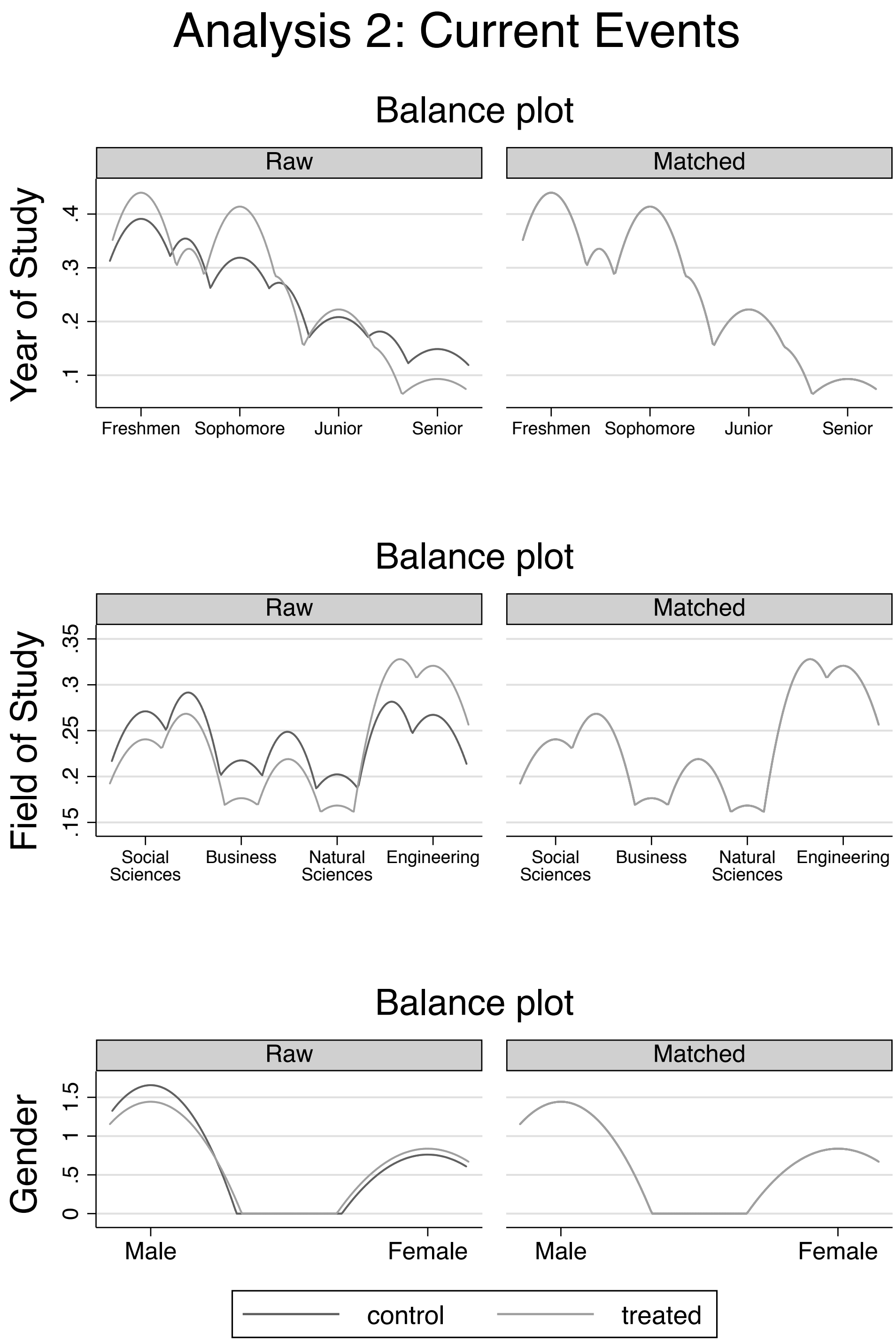


Figure K: Matching performance of current event model.

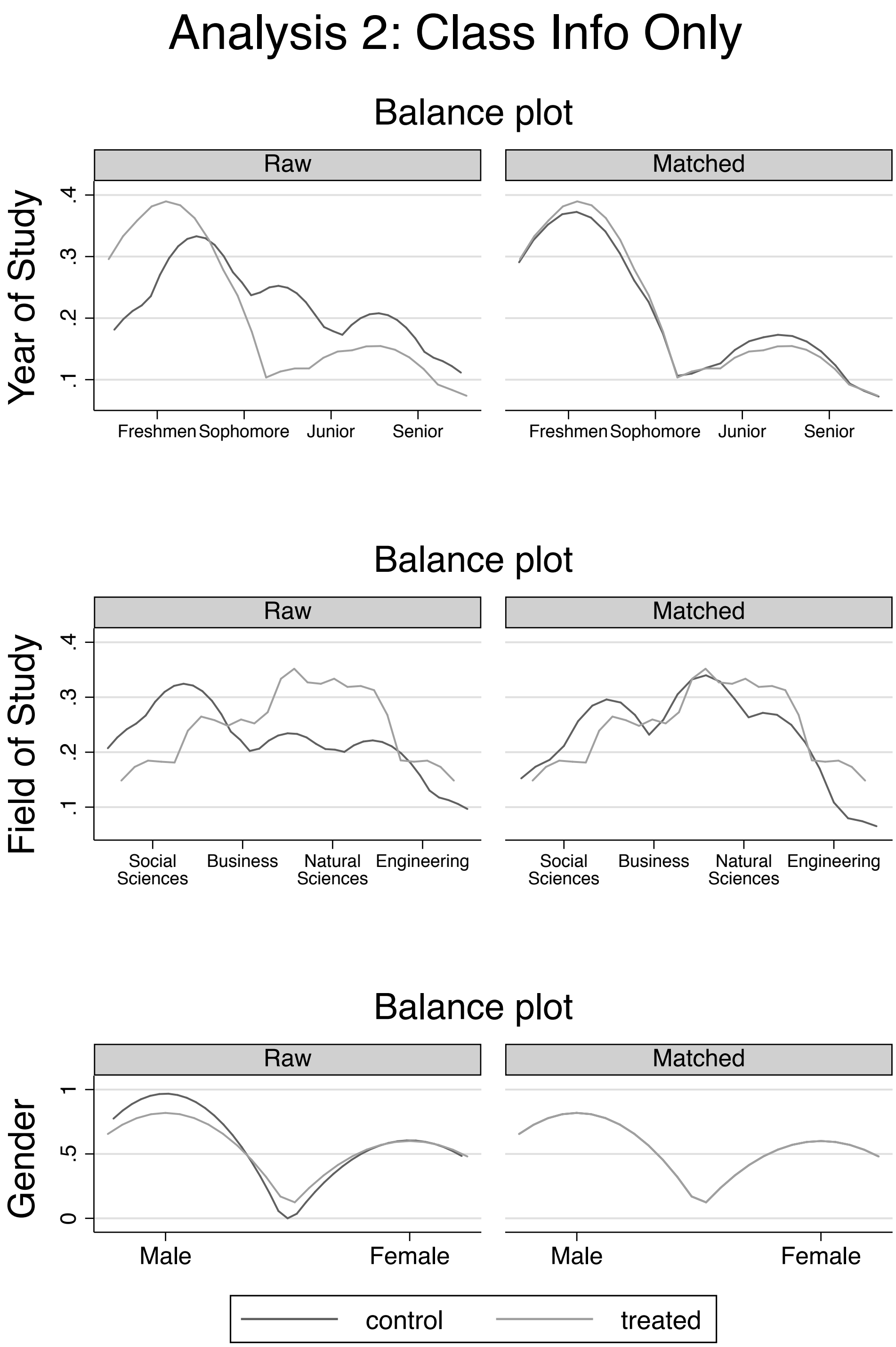


Figure L: Matching performance of class information model.

4.3 Analysis 3

Describe model again: Comparison to students that received class information only

The third analysis compares students that received one of the three treatments (skill, topics, current events) to students that received class information only. The balance plots below compare the treatment and the control groups for each treatment across the three covariates used for matching. Covariance balance is improved by matching, though performance is not quite as good as with Analysis 1 and 2 due to a lower number of observations available for matching.

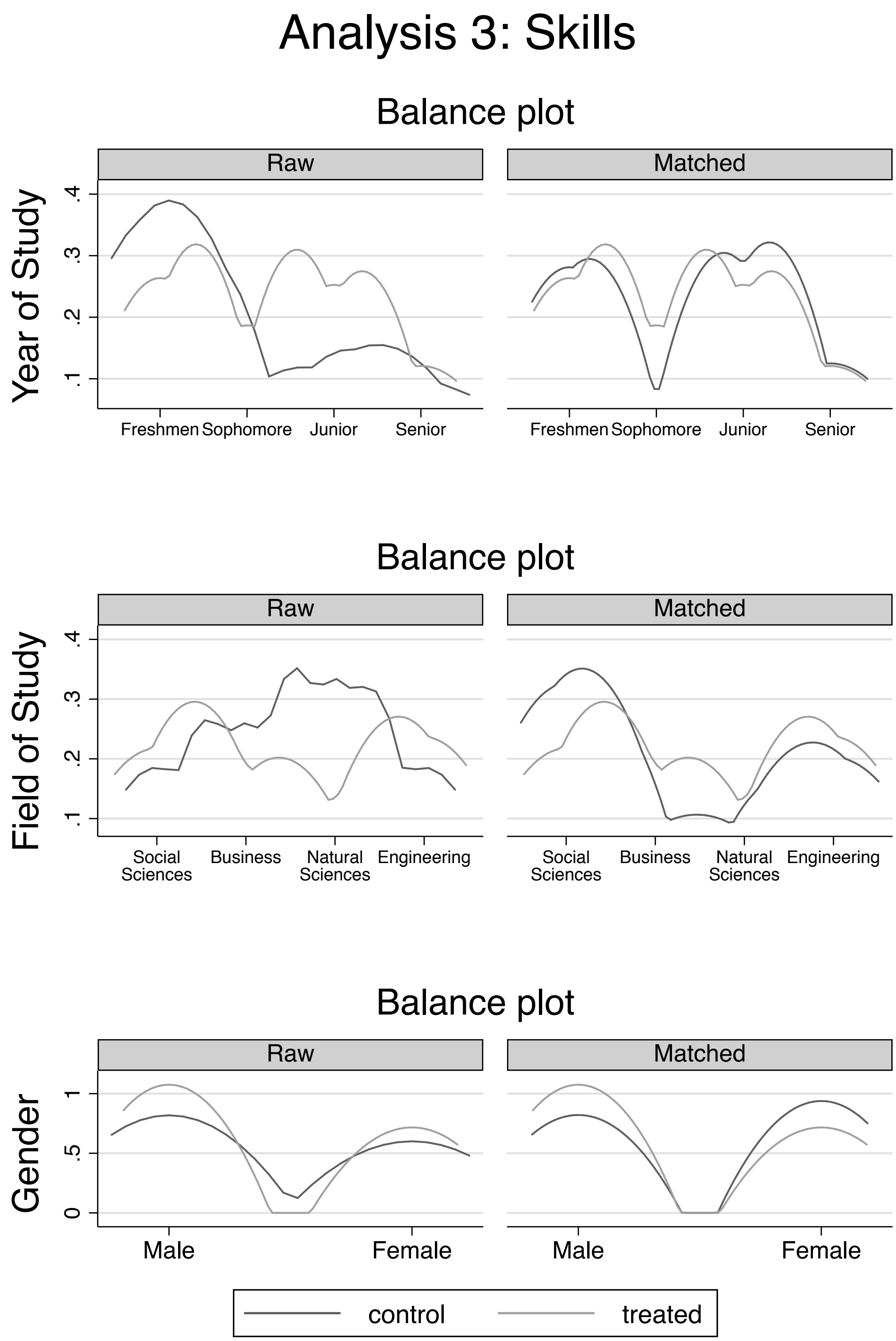


Figure M: Matching performance of skill model.

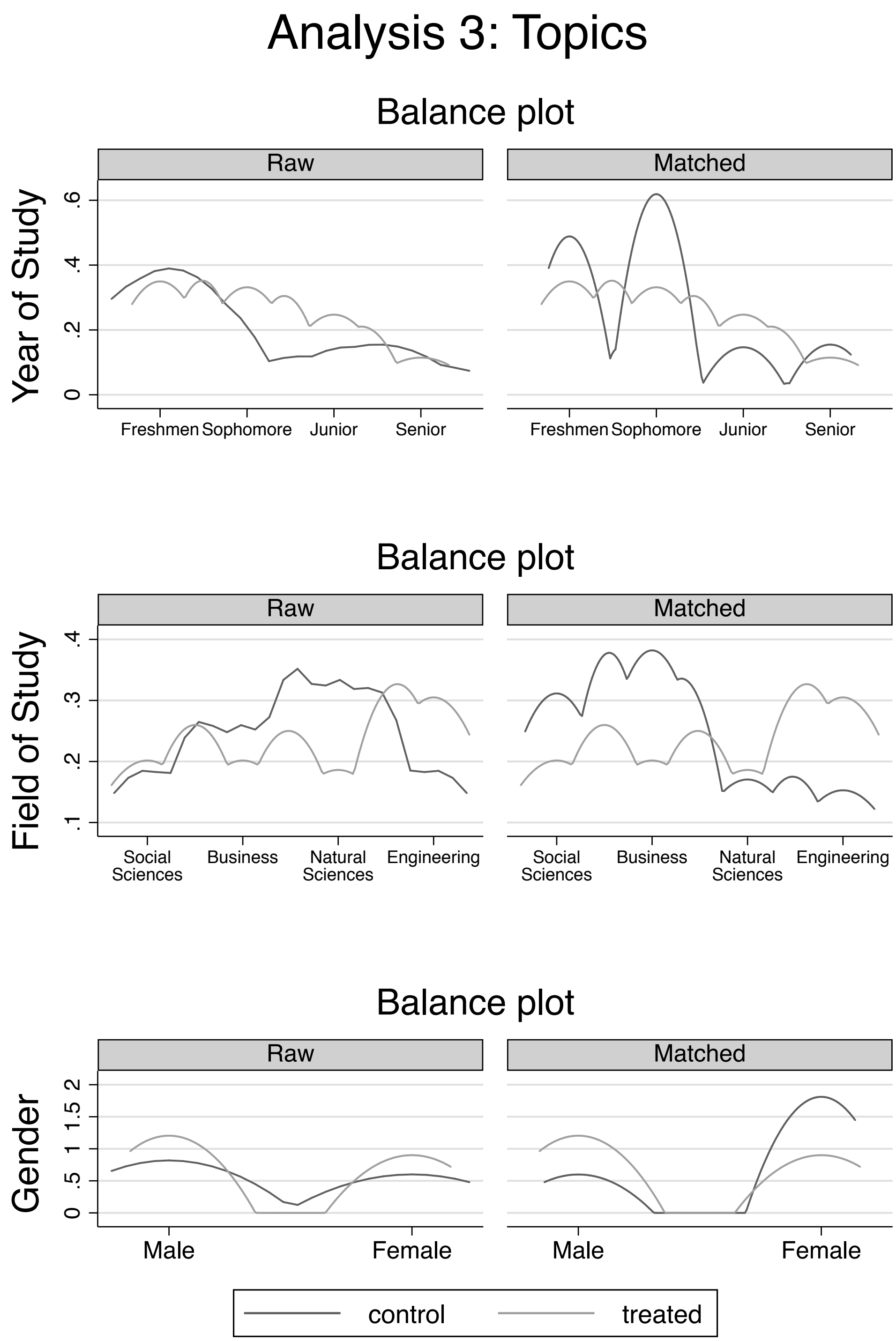


Figure N: Matching performance of topics model.

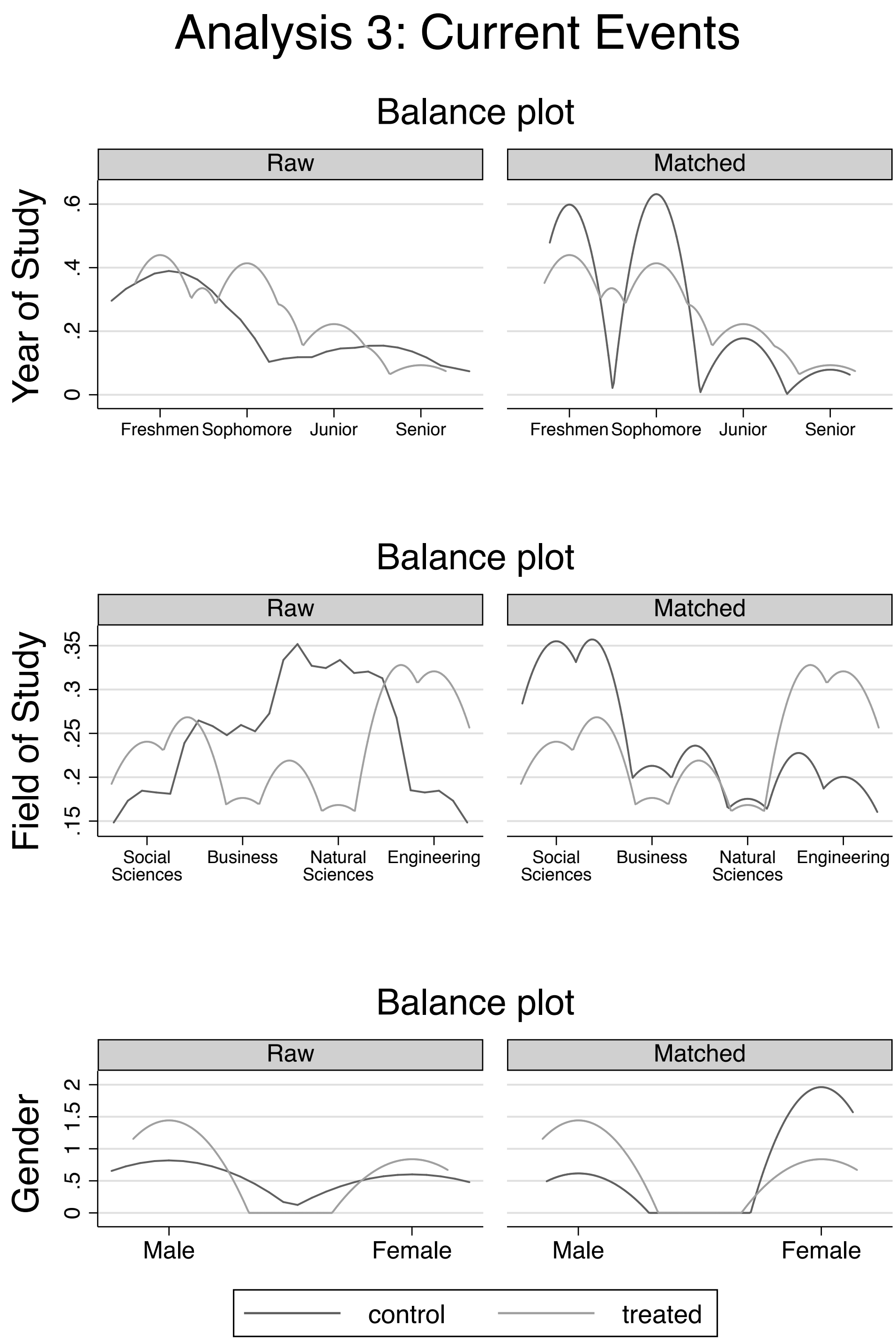


Figure O: Matching performance of current event model.