Online supplementary material

A. Possible selection effects on missing date of first romantic relationship

To check possible selection effects on missing data, we performed a logistic regression analysis where the response variable was if the observation had (Y=1) or not (Y=0) a missing value on the date for the first romantic relationship. We included a wide series of covariates in such a way to identify possible associations between individual characteristics and missingness through a backward stepwise procedure. The final model presented in Table 1.A comprised all the covariates that proved to be significant or weakly significant, and more specifically: student's gender, student's birth cohort, the age at which the student had a first sexual intercourse (in classes), student's final lower secondary education score, mother's work when the respondent was 11-13 years old, father's birth cohort, if the parents were married, cohabiting or separated when the student was 11, the municipality size and the macro-area of residence during high school. All variables about life experiences (see Section 5.2 for a list of these covariates), the age at which the student had a first incomplete sexual experience (in classes), the type of high school attended and other family background covariates (such as the highest level of education among the two parents), proved not to be significant.

Table 1.A shows that the pattern of missingness was associated with student's gender and birth cohort, with a higher non-response among men and among youngest cohorts; students who did not have a first intercourse had a lower probability of a missing value on the date for the first romantic relationship; the missingness is also negatively associated with the final mark at lower secondary; students living in municipalities between 10,000 and 50,000 inhabitants during high school had a (weakly) lower probability of a missing value; and finally, those living in the Centre and mainly in the South of Italy had a higher probability of a missing value.

In addition, many variables included in the model had an unknown category, such as mother's work, father's birth cohort, parental marital status, the municipality size and the macro-area of residence during high school. All the coefficients of the unknown categories of the corresponding covariates are significant (weakly significant in one case); thus, it emerges an association between the unknown category and the response variable, namely a missing value on the date for the first romantic relationship.

To sum up, the missing value on the date for the first romantic relationship does not seem to be associated with sexual life events, given that only those not having had a sexual intercourse had a lower probability of a missing value. Possible selection effects on missing value involve, instead, students possessing the above-mentioned characteristics, such as being men, belonging to the

youngest cohorts and having had a low final score at lower secondary school. Finally, it emerges that students who not answered accurately to the date of first romantic relationship did the same for other questions in the survey.

Table 1.A: Model coefficients from logistic regression model on the presence of a missing value on the date of first romantic relationship.

	coef	std error	p-value
Gender (ref. Male)			
Female	-0.378	0.075	0.000
Birth cohort (ref. 1975-1979)			
1980-1981	-0.126	0.152	0.407
1991-1995	0.627	0.156	0.000
1996-1997	0.533	0.156	0.001
Age at first intercourse (ref. <15)			
15-17	-0.203	0.141	0.150
17-19	-0.126	0.137	0.357
>19	-0.233	0.148	0.116
no intercourse	-1.225	0.168	0.000
Final score at lower secondary	-0.143	0.032	0.000
Mother's work (ref. No)			
Yes	-0.054	0.080	0.500
Unknown	0.564	0.164	0.001
Father's birth cohort (ref. 1919-1945)			
1946-1955	-0.098	0.157	0.531
1956-1965	-0.029	0.181	0.875
1966-1980	0.009	0.195	0.964
Unknown	0.622	0.191	0.001
Parental marital status (ref. Married parents)			
Cohabiting parents	0.237	0.259	0.361
Separated parents	0.116	0.169	0.494
Unknown	0.831	0.178	0.000
Municipality size during high school (ref. <10,000 inhab)			
10,000-50,000 inhab	-0.172	0.095	0.071
50,000-100,000 inhab	-0.080	0.116	0.488
100,000-500,000 inhab	-0.202	0.129	0.117
>500,000 inhab	-0.046	0.122	0.704
Unknown	0.333	0.178	0.062
Macro-area of residence during high school (ref. North)			
Centre	0.275	0.106	0.010
South	0.835	0.088	0.000
Abroad	0.449	0.375	0.232

Unknown	1.324	0.302	0.000
Constant	-0.781	0.329	0.018

B. Analyses on the reduced sample

As summarised in Section 6.3 in the main text, we repeated our analyses on the reduced sample of 7,262 students. We reconstructed the sequences for this sample and then conducted the sequence analysis with the OM algorithm and cluster analysis using Ward's algorithm, following the same criteria as described in Section 5.1 in the main text.

Figure 1.B shows the state distribution plots of the solution given by six clusters, which can be compared with Figure 2 in Section 6.1 in the main text. Graphically, the six clusters appear very similar in the imputed sample and in the reduced sample. Nevertheless, some differences may be detected between the two analyses. First, the numerosity of each cluster differs (see Table 1.B below, which is identical to Table 3 in the main text); while in some cases the difference is restrained (e.g., for the sex without commitment cluster), for others the gap is wide (i.e. for the forerunners profile). Second, median ages at first sexual intercourse slightly change in the two groups of clusters, being lower in the reduced sample than in the imputed sample (except in the romantic love cluster where the difference is negligible).

Figure 1.B: State distribution plots of clusters of students' affective and sexual life courses. Reduced sample.

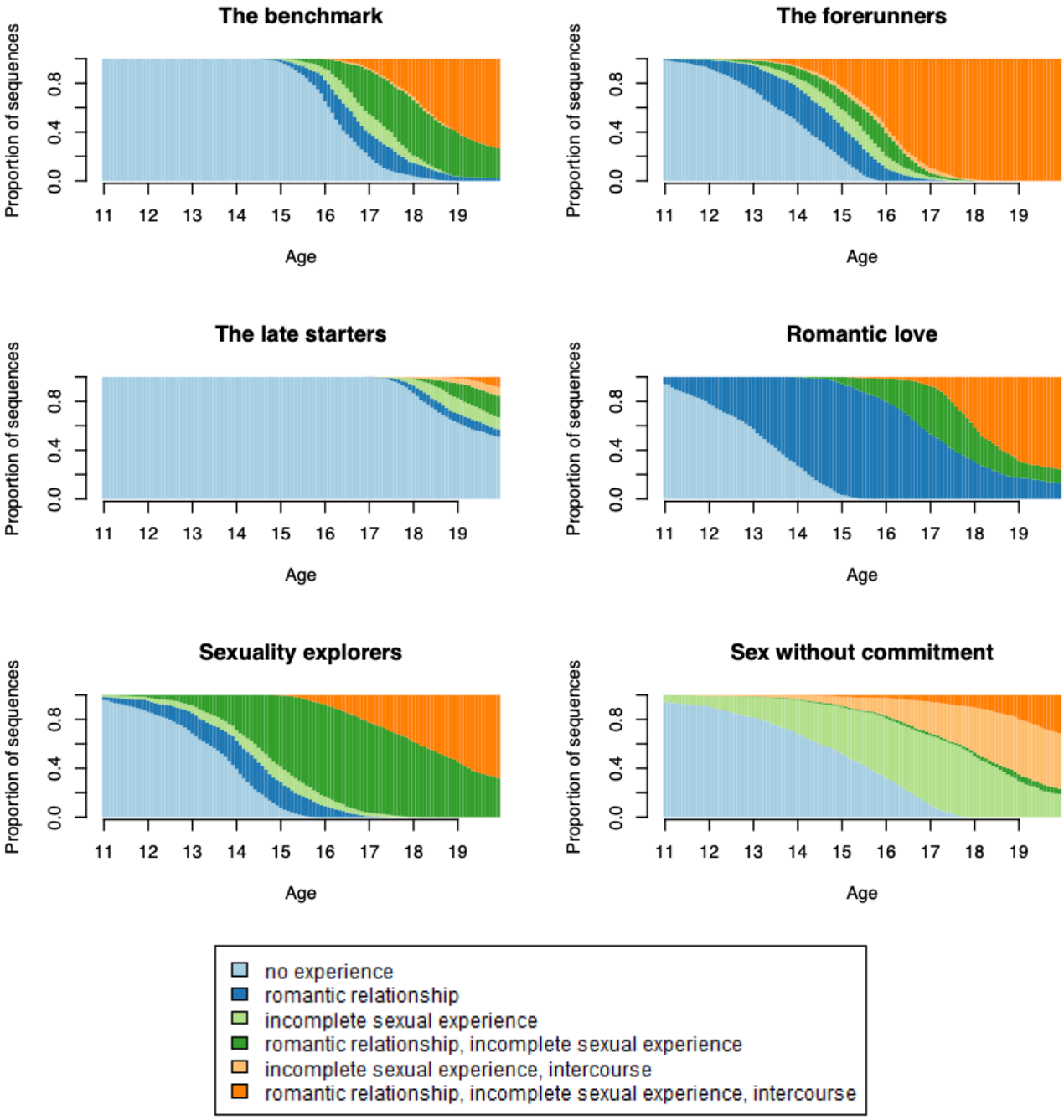


Table 1.B: Clusters of students' affective and sexual life courses. Students' absolute and percentage frequencies (by column) in the original, imputed sample (8,243 students) and in the reduced sample (7,262 students)

	Original sample		Reduced sample	
Cluster label	abs.v.	*%*	abs.v.	*%*
The benchmark	1,380	*16.7*	1,536	*21.2*
The forerunners	2,687	*32.6*	1,265	*17.4*
The late starters	757	*9.2*	1,172	*16.1*
Romantic love	1,186	*14.4*	665	*9.2*
Sexuality explorers	1,237	*15.0*	1,565	*21.6*
Sex without commitment	996	*12.1*	1,059	*14.6*
Total	**8,243**	***100.0***	**7,262**	***100.0***

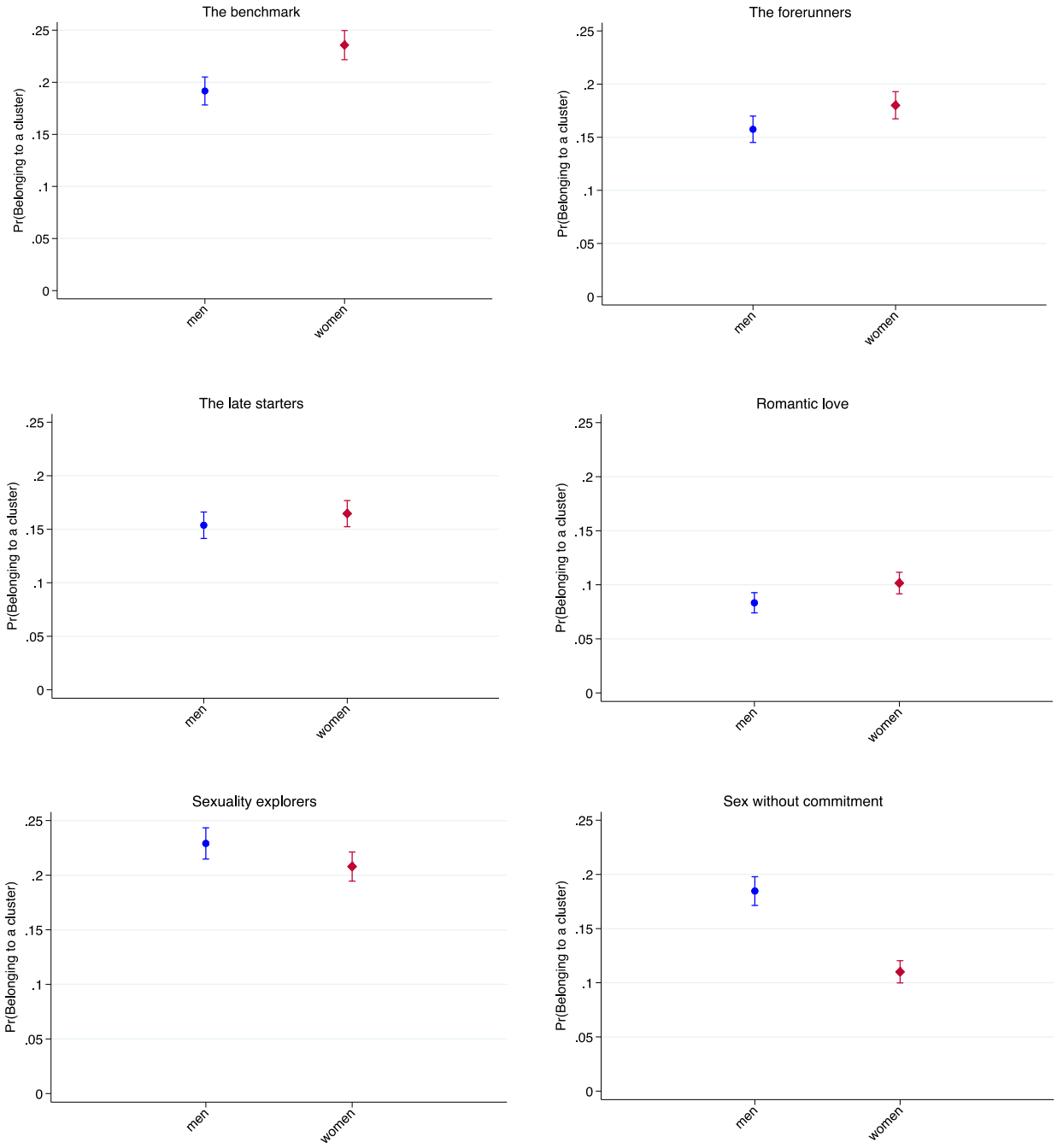
Figure 2.B shows confidence intervals of predicted probabilities of belonging to the six clusters according to student gender (Figure 2.B1) and birth cohort (Figure 2.B2) deriving from Model *E* (complete model, without the interaction term between gender and birth cohort) for the reduced sample.

Specifically, we see in Figure 2.B1 that the male and female confidence intervals of the predicted probabilities do overlap in four clusters: the forerunners, the late starters, romantic love and sexuality explorers. With respect to the results obtained for the imputed sample, gender differences appear attenuated, thus suggesting a more marked gender equality. The main noteworthy difference is about the forerunners, whereas the others three clusters were overlapping or very close also in the imputed sample.

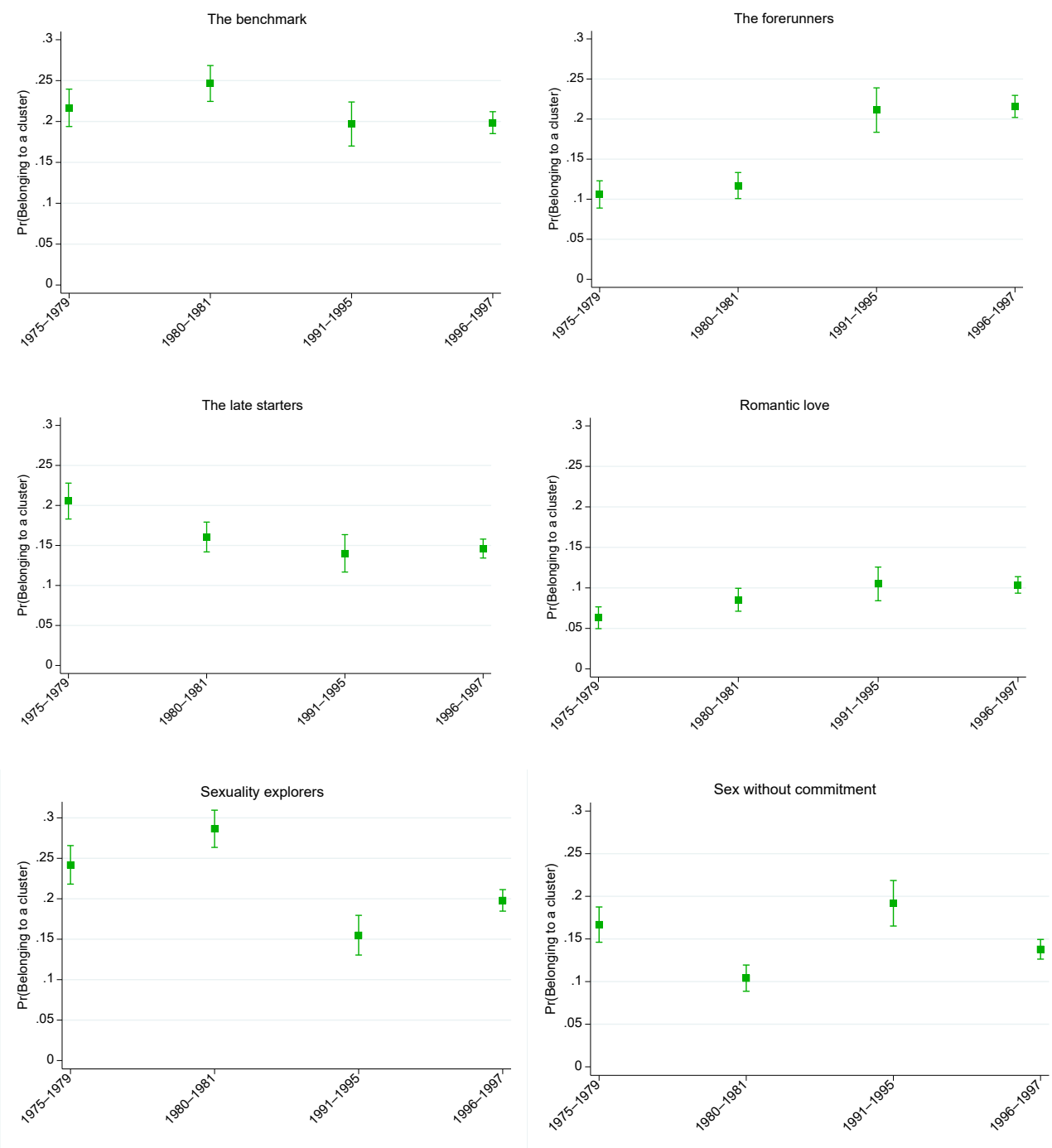
When looking at the differences by birth cohort (Figure 2.B2), the patterns of the six clusters are very similar comparing the imputed sample and the reduced sample, thus confirming the trend of the various pathways over time.

Figure 2.B: Results from Model *E*: Predicted probabilities of belonging to the six clusters according to student gender (Fig. 2.B1) and birth cohort (Fig. 2.B2). CI 95%. Reduced sample.

1) By gender



2) By birth cohort

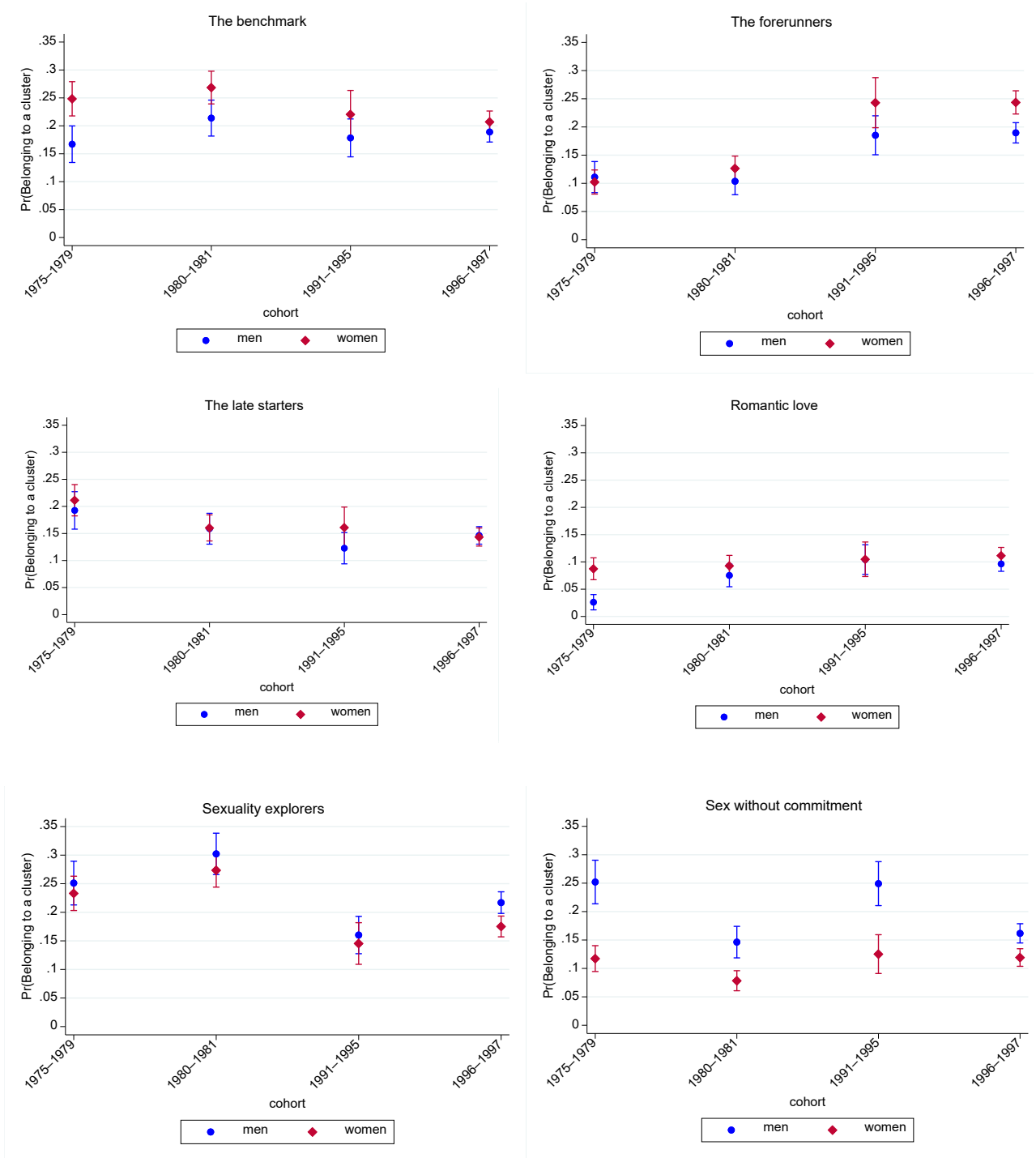


Note: own elaboration on SIS and SELFY data. To estimate predicted probabilities, gender (in Figure B, a) and birth cohort (in Figure B, b) are allowed to vary, while gender (in Figure B, b), birth cohort (in Figure B, a), student's final lower-secondary score, type of high school attended, frequency of church attendance when student was 13, smoking when student was 11 to 13, attendance of discotheque when student was 11 to 13, parental highest level of education, mother's work when respondent was 11 to 13, parental partnership when student was 11, and macro-area of residence during high school are kept at the mean value.

Figure 3.B shows confidence intervals of predicted probabilities of belonging to the six clusters according to student gender and birth cohort deriving from Model *F* (complete model, with the interaction term between gender and birth cohort) for the reduced sample.

On one hand, comparing Figure 3.B with Figure 4 in the main manuscript, we can see that two clusters, namely the late starters and romantic love, have the same pattern and overlapping confidence intervals. On the other hand, the forerunners is again the cluster with the most relevant differences; clearly, this group is different in the two solutions. The remaining three clusters, namely the benchmark, sexuality explorers and sex without commitments, show some slight differences in the overlapping confidence intervals, but the trends are superimposable.

Figure 3.B: Results from Model *F*: Predicted probabilities of belonging to the six clusters according to student gender and birth cohort. CI 95%. Reduced sample.



Note: own elaboration on SIS and SELFY data. To estimate predicted probabilities, gender and birth cohort are allowed to vary, while student's final lower-secondary education score, type of high school attended, frequency of church attendance when student was 13, smoking when student was 11 to 13, attendance of discotheque when student was 11 to 13, parental highest level of education, mother's work when respondent was 11 to 13 years old, parental partnership when student was 11, and macro-area of residence during high school are kept at the mean value.