Express $\frac{81}{108}$ as a decimal.

What do student exam results tell us about free education in PNG?

Anthony Swan
Development Policy Centre
9 November 2015

Outline

- Research questions
- Education in PNG
- TFF
- Grade 8 exams
- Multilevel modelling
- Descriptive statistics
- Model
  - Approach
  - Identification strategy
- Results
- Policy implications

Broad research questions

- What are the relationships between student and school characteristics and student learning outcomes?
- How has the free education impacted on student learning outcomes?
- What can education reformers learn from this experience and what are appropriate policy responses going forward?

More specifically…

- How do greater student numbers impact on test scores?
- What is the relationship between student age and test scores?
- What is the effect of higher average class age on test scores at the individual level?
- What effect does gender have on test scores depending on location, interactions with female teachers and other students of the same gender?
- How does remoteness impact on test scores?
- How much of the variation in test-scores is due to factors not directly related to individual students and how is it changing over time?
- How are relationships between student and school characteristics changing over time and their relationship with free education policy?

State of education in PNG

- Key indicators
  - Gross enrolment for basic education was 71% in 2000, 96% in 2014, and for secondary education it was 12.5% in 2000 and 22.5% in 2013.
  - Around 7% average annual growth in enrolment
  - Transition rate from gr8 to gr9 about 50% for girls and 60% for boys in 2014
  - Challenges: expand the system capacity in line with popn. growth, the remoteness of communities, lack of resources, poor coordination between national and sub-national levels, complexity of society, lack of teachers, substandard curriculum (Education plan, draft).

Education policies in PNG

- The education system
- Free education attempts
- Outcomes based education
- School BOM
- SLIP
- Govt funding allocations
- TFF from 2012
- New National Education Plan (in draft form)
  - Focus on quality education (?)
  - OBE to be abolished and a standards based curriculum to be implemented
  - National grade 8 and 10 exams to be abolished

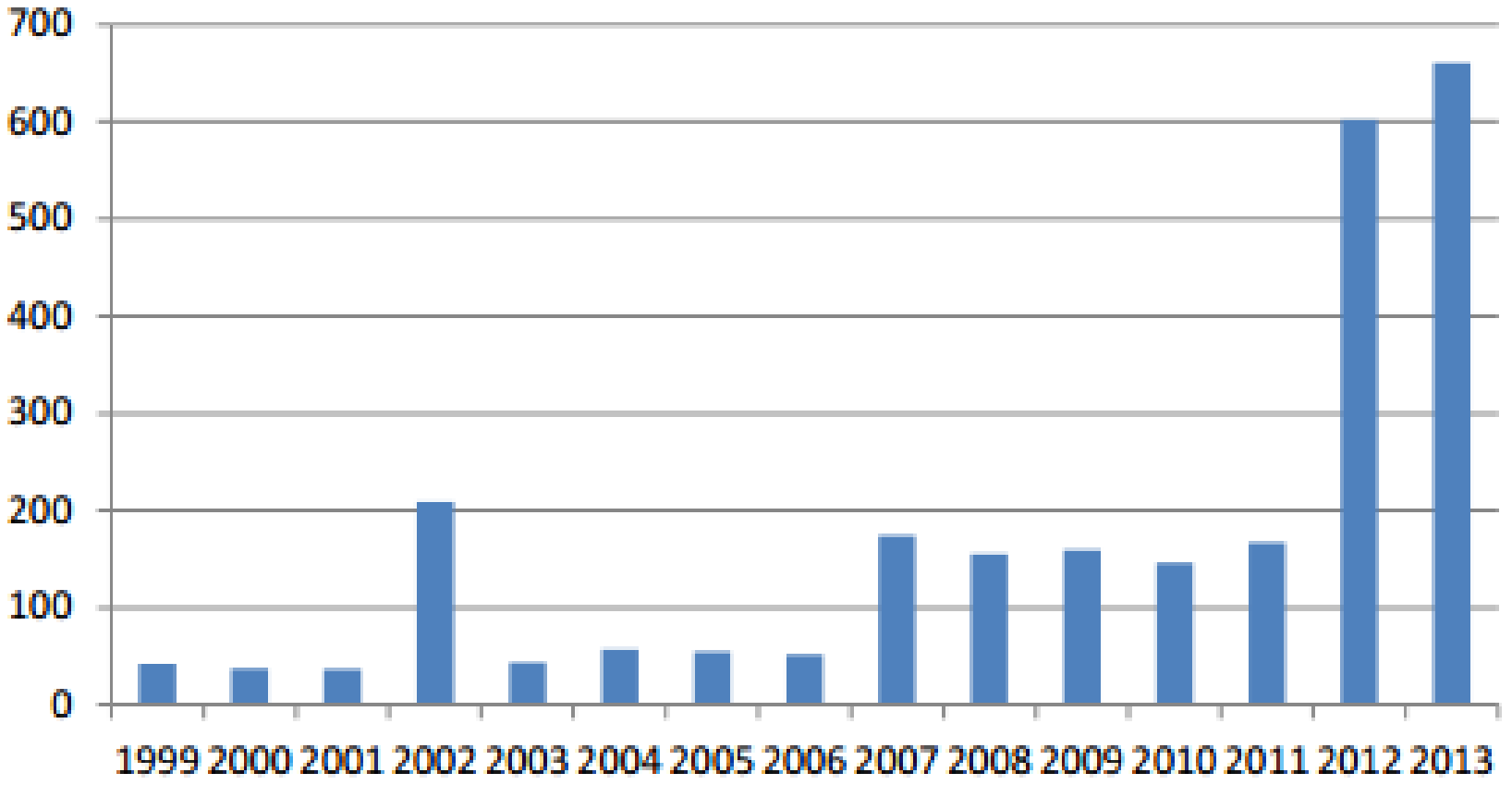
Tuition Fee Free Policy

- In 2012, abolished school fees in basic education and subsidised secondary education (75%)
- Aim:

1. Access is improved for all children, especially for girls;
2. Retention is enhanced where more children complete nine years of primary education;
3. Quality of education is improved for all grades of elementary and primary levels;
4. Education management is strengthened across all administrative levels; and
5. Equity is enhanced to ensure quality education is available for all children in all communities across the country.

- Direct subsidy payments to schools – K270 per enrolled student
- Big bang approach to implementation
  - 15.4% increase in enrolment in 2012 (4.1% in 2013)
  - can cause “access shock”, whereby a sudden rise in student numbers puts pressure on educational quality.
  - School Boards are key decision makers on how subsidies are spent.

Subsidy payments from the central government (2012 prices, million kina)



Grade 8 Primary School exams

- What are they?
  - English, Written Expression (ELP), Maths, Combined Subjects.
  - Assessment of learning outcomes based on the OBE curriculum. Fairly consistent over time.
  - Marked at the province level
    - Possible variation in grading over location and time.
- What are they used for?
  - A way to select students into secondary school.
  - Make comparisons across areas in PNG and gender
  - A tool to improve teaching and learning?

A multilevel modelling approach

- Let's start with a couple of simple models:

(a) $$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + R_{ij}$$

(b) $$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + \beta_3 x_{ij} z_j + R_{ij}$$

where Y is the test score, β_1 is the coefficient on the student level variable X, β_2 is the coefficient for the school level or group level variable Z, and β_3 is the coefficient on the cross level interaction variable XZ. The variable R_{ij} is the residual or error. We will assume for now that all residuals are mutually independent and normally distributed with mean zero and constant variance across groups.

- Is there anything wrong with this? Yes, if the data has a meaningful multilevel structure: e.g. if the effect of different student and school characteristics on test scores vary across schools.

Multilevel modelling

- We can account for variation across groups by allowing the coefficients (intercept and slopes) to depend on the group:

$$Y_{ij} = \beta_{0\boldsymbol{j}} + \beta_{1\boldsymbol{j}} x_{ij} + \beta_2 z_j + R_{ij}$$

- Let's start with the variable intercepts model (i.e. slope coefficients are assumed identical across groups) and no Z variables for simplicity. We model the varying intercept at the group level:

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

and substituting back in gives:

=> $$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + R_{ij}$$

where $\gamma_{10} \coloneqq \beta_{1\boldsymbol{j}}$ (i.e. just a change in notation).

- The variable intercepts model can be thought of as one where the U_{0j} are fixed parameters such as the effect of gender. It is not possible to include group level variables Z with this specification because they would be perfectly collinear with the fixed group effects.
- Alternatively it can be specified as a random effects model which is the approach here.

Random intercept model

- The random effects model assumes the U_{0j} are (iid) random variables so they are really unexplained group effects or group residuals, controlling for the effects of variable X. We assume the U_{0j} are randomly drawn from a population with zero mean and unknown variance. There is only one parameter associated with the U_{0j} which is its variance.
- In the random effects model it is possible to explain test score with group level variables as well as the U_{0j}:

$$Y_{ij} = \gamma_{00} + \gamma_{10}x_{1ij} + \gamma_{20}x_{2ij} + \cdots$$
$$+\gamma_{01}z_{1j} + \gamma_{02}z_{2j} + \cdots + U_{0j} + R_{ij}$$

Random Slope model

- Slopes can also be random. Suppose:

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + R_{ij}$$

- Then we can model the random intercept and slope as:

$$\beta_{0j} = \gamma_{00} + U_{0j}, \qquad \beta_{1j} = \gamma_{10} + U_{1j}$$

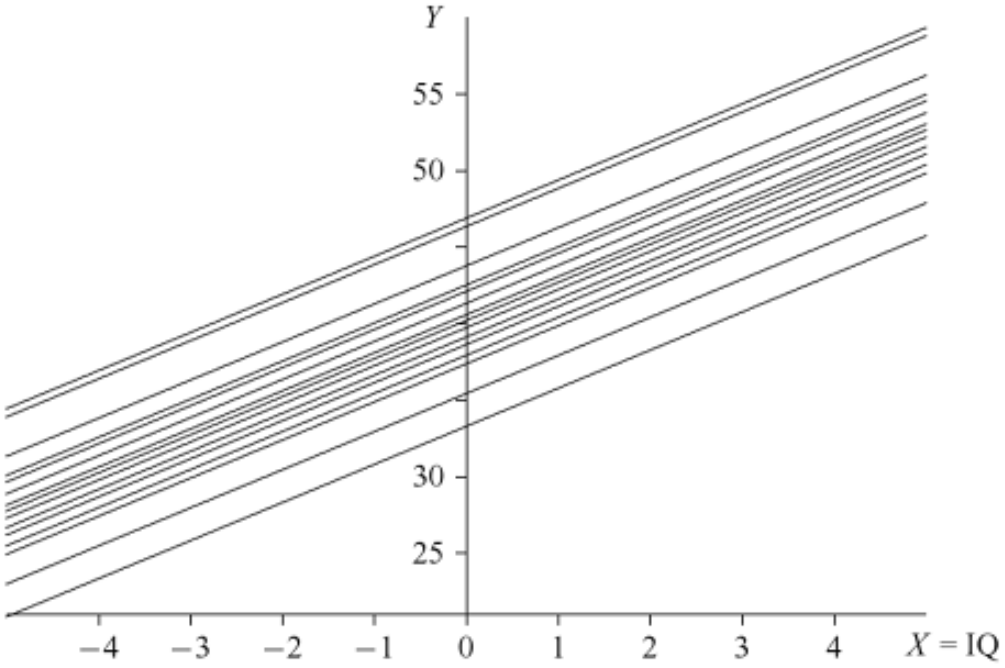
and substituting back in gives:

=> $$Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + U_{1j}x_{ij} + R_{ij}$$

Average intercept

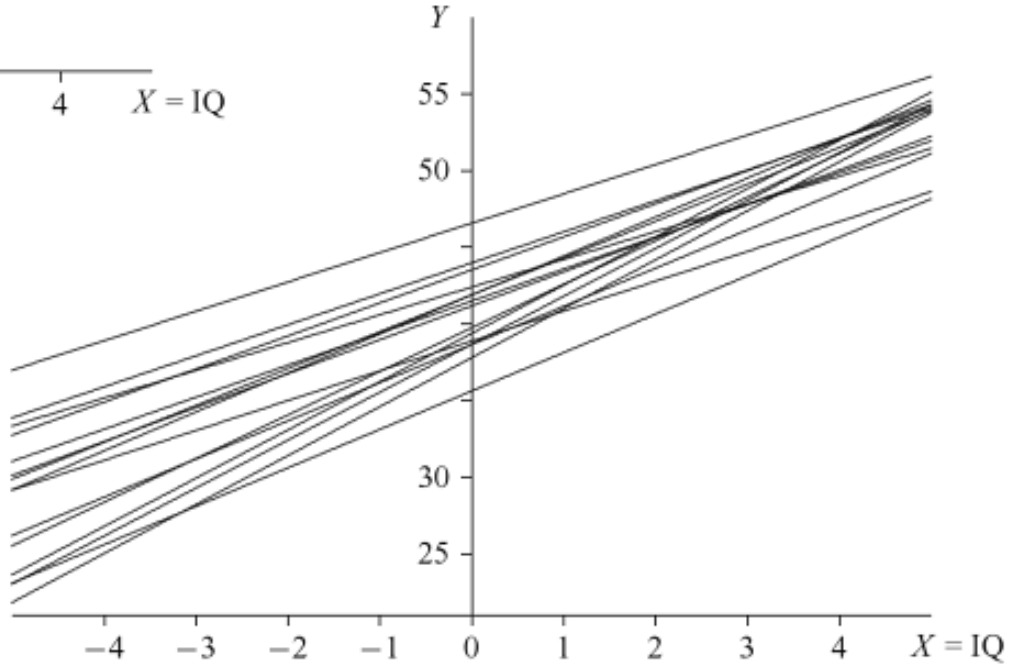
Average slope

- The first 2 terms on the RHS form the fixed part of the model and the last 3 terms form the random part of the model.
- U_{0j} and U_{1j} will usually be correlated within the group. However we assume that the pair of effects (U_{0j} , U_{1j}) are independent and identically distributed across groups as well as the individual level residuals R_{ij}.



Random intercepts

Random intercepts and slopes



DESCRIPTIVE STATISTICS

	2011			2012			2013		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
English	55964	44%	16%	65016	46%	19%	69835	52%	18%
ELP	55882	24%	14%	64667	27%	15%	69800	25%	14%
Maths	55950	39%	17%	64989	39%	19%	69849	40%	19%
Combined	55877	52%	14%	65004	49%	16%	69798	49%	15%
Female	56636	41.6%	49.3%	65742	41.3%	49.2%	70697	41.2%	49.2%
Age	56636	16.8	1.9	65742	17.0	2.1	70697	17.1	2.1
School Female Share	1699	41%	14%	1845	41%	14%	1875	41%	14%
School Age Mean	1699	17.2	1.3	1845	17.4	1.4	1875	17.5	1.4
Govt. School	1699	44%	50%	1845	44%	50%	1875	43%	49%
Total G8 Students per sch.	1699	33.7	29.1	1845	35.8	30.3	1875	38.0	34.9
Enrolled G8 per sch.	1699	41.3	33.4	1845	43.0	35.3	1875	44.9	37.2
Attendance (ENG)	1699	82%	20%	1845	83%	22%	1875	84%	29%
Attendance (ELP)	1699	82%	20%	1845	82%	22%	1875	84%	29%
Attendance (MATHS)	1699	82%	20%	1845	83%	22%	1875	84%	29%
Attendance (COMB)	1699	82%	20%	1845	83%	22%	1875	84%	29%
Female Attendance	1698	82%	26%	1840	82%	28%	1874	83%	32%
Male Attendance	1699	85%	22%	1845	86%	25%	1875	89%	35%
Student Teacher ratio	1699	21.6	11.3	1845	22.4	10.9	1875	23.5	14.5
Teachers per sch.	1699	1.72	1.34	1845	1.74	1.34	1875	1.75	1.35
Male Teachers per sch.	1699	1.12	0.92	1845	1.11	0.91	1875	1.12	0.95
Female Teachers per sch.	1699	0.60	0.96	1845	0.62	0.95	1875	0.63	0.96
Female Teacher Share	1699	30%	40%	1845	31%	40%	1875	33%	41%
Remoteness (PARI)	1699	0.86	0.85	1845	0.84	0.82	1875	0.90	0.89

MODEL

Model specification

- Selection of relevant explanatory variables (and interactions) in the fixed part and random slopes at various nested levels in the random part.
- Used a general approach by selecting relevant level 1 and level 2 and level 3 variables and including plausible (cross-level) interactions.
- Student age and gender were chosen to have random slopes, as well as year variable.
  - Random slopes were justified based on LR tests
- Attendance was excluded from the model due to likely endogeneity.

What levels explain variation in test scores?

Intraclass Correlation Coefficients (Unconditional):

	2011		2012		2013	
	School	LLG	School	LLG	School	LLG
English	0.261	0.104	0.361	0.192	0.339	0.161
Maths	0.404	0.231	0.399	0.243	0.406	0.230
Combined	0.262	0.102	0.314	0.156	0.326	0.164
ELP	0.310	0.168	0.320	0.173	0.386	0.254

- ICCs are increasing or at least not decreasing: non-student level factors seem to matter more over time.
- Rule-of-thumb: include a level in a multilevel model if ICC>0.1

  => Include LLG level but not Province level

Fixed part of the model

- Standardised test-score modelled as a linear function of:

  age (class-mean-centred); gender;

  total G8 students (grand-mean-centred);

  female teacher share;

  female student share (grand-mean-centred);

  mean class age (grand-mean-centred);

  agency; student-teacher-ratio (grand-mean-centred); remoteness; and year;

- and interactions of:
  - female_teacher_share#female_student_share#gender
  - age#mean_class_age#gender
  - gender#student_teacher_ratio#agency
  - year (with all variables)

Random part of the model

- Standardised test-score modelled as a linear function of:
  - Random intercepts at the school and LLG level
  - Random slope at the school level consisting of:
    - age, gender, and year dummy
  - Random slope for year dummy at the LLG level.
  - Unstructured variance and covariances of random terms

Identification strategy

- Data is limited at the individual, class and school level
  - Missing socioeconomic status, IQ, parental education, household characteristics…
  - Missing teacher characteristics, time spent teaching…
  - Missing school infrastructure level, teaching resources, community engagement, school oversight…
- Causal interpretation to estimated effects is a problem when observed data is correlated with unobserved data.
- The data we have are proxies for underlying characteristics so a good deal of interpretation and caution is needed in the analysis.
  - e.g. student age is a proxy for SES and IQ

Identification strategy

- How to interpret and treat variation in test score means and variance over time?
  - Here I have standardised test scores (mean 0, variance 1) over time and at the national level to account for differences in difficulty of exams and grading.
  - Downside: absolute effects of changes over time are eliminated (apart from year level random effects) => focus on relative effects over time.
  - The effect of the TFF is assumed to be associated with the difference over time (i.e. 2012 & 2013 relative to 2011) in the (average) marginal effect of particular variables of interest.

About margins and marginal effects.

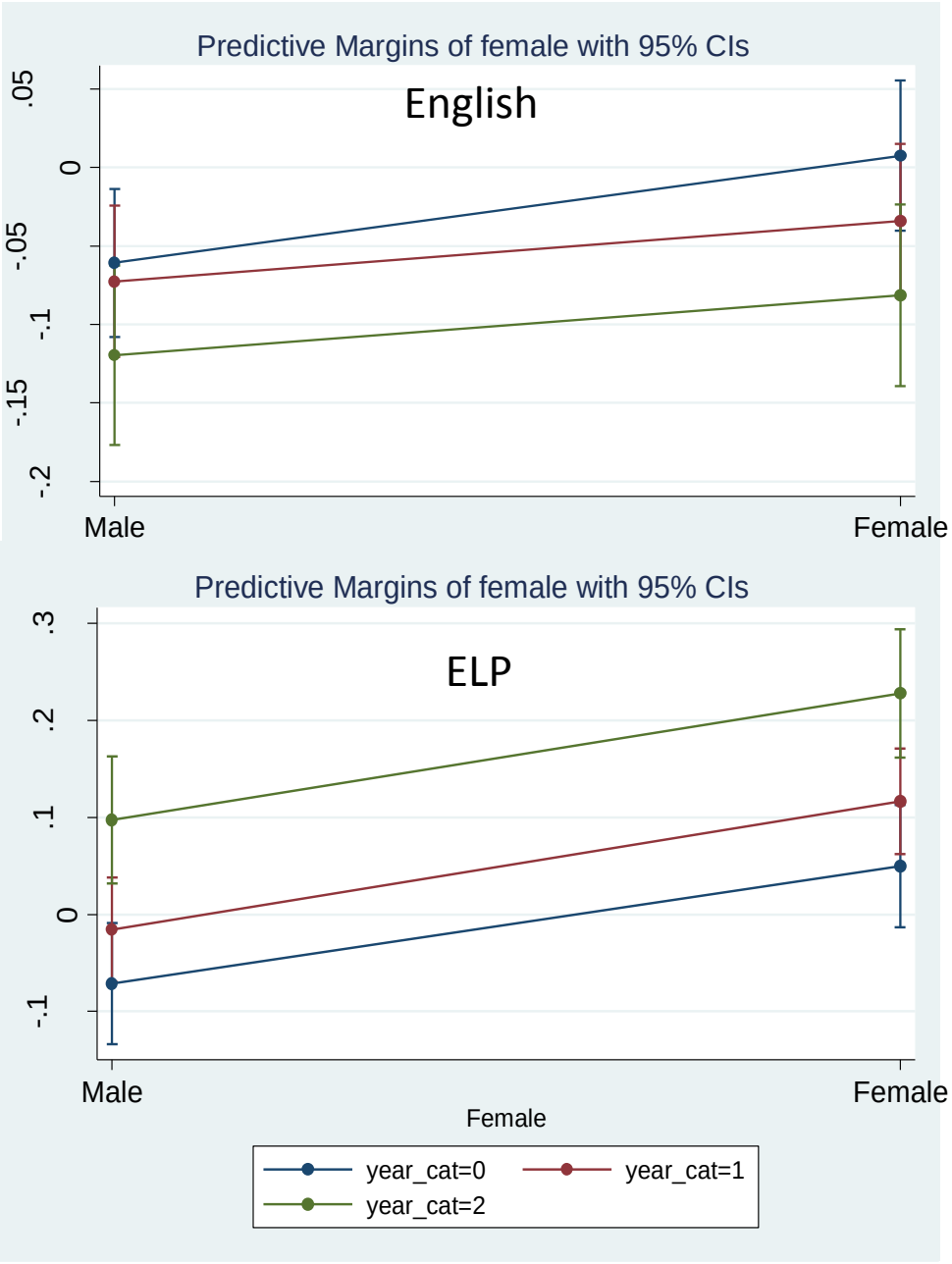
- A margin is a statistic based on a fitted model calculated over a dataset in which some of or all the covariates are fixed at values different from what they really are.
- For instance, after a linear regression fit on males and females, the marginal mean (margin of mean) for males is the predicted mean of the dependent variable, where every observation is treated as if it represents a male; thus those observations that in fact do represent males are included, as well as those observations that represent females. The values of all other variables are held constant.
- A marginal effect is the change in the margin when a variable is fixed at a value and then at a different value (e.g. a unit increase for a continuous variable, or going from male to female for a binary variable).

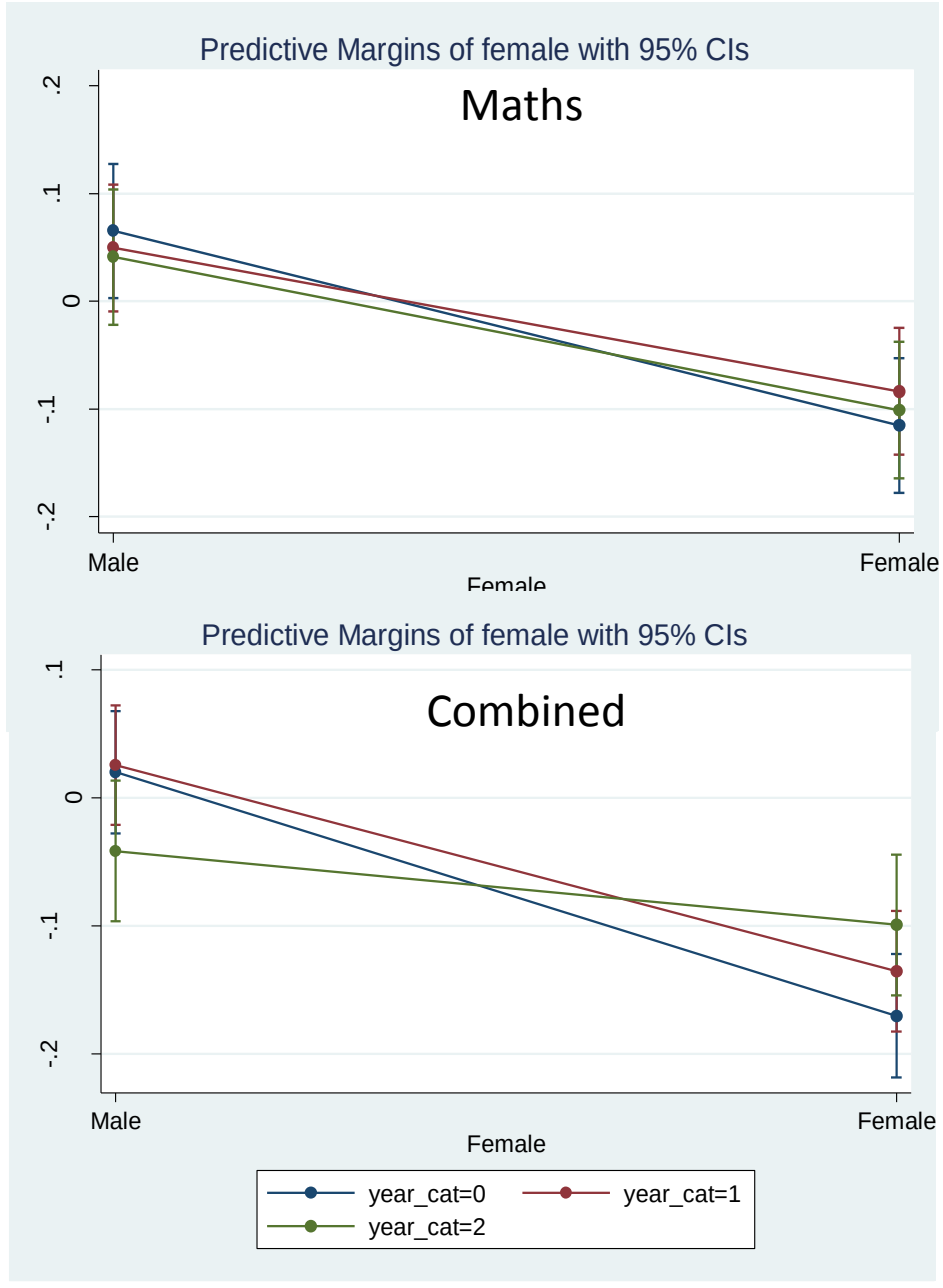
RESULTS

Results

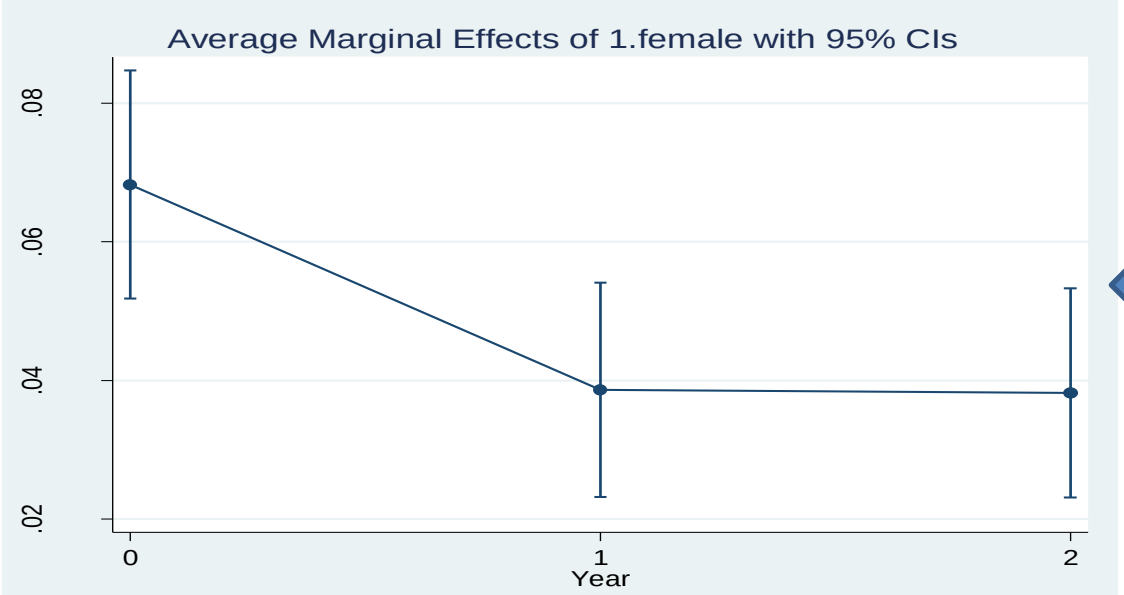
GENDER

Overall female effect

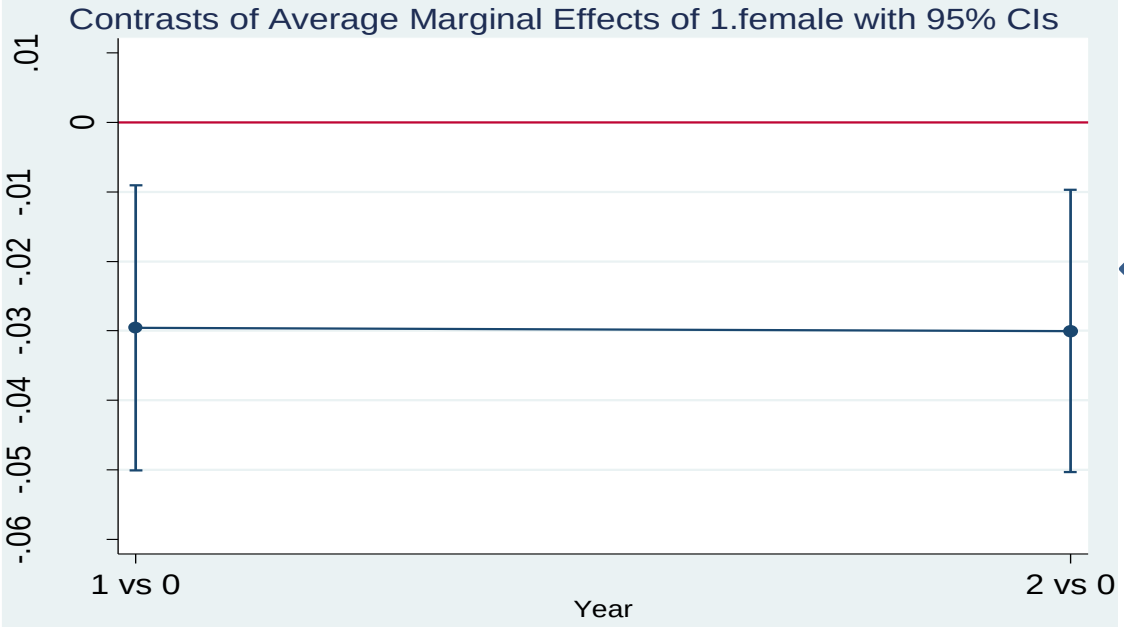




Overall female effect - English

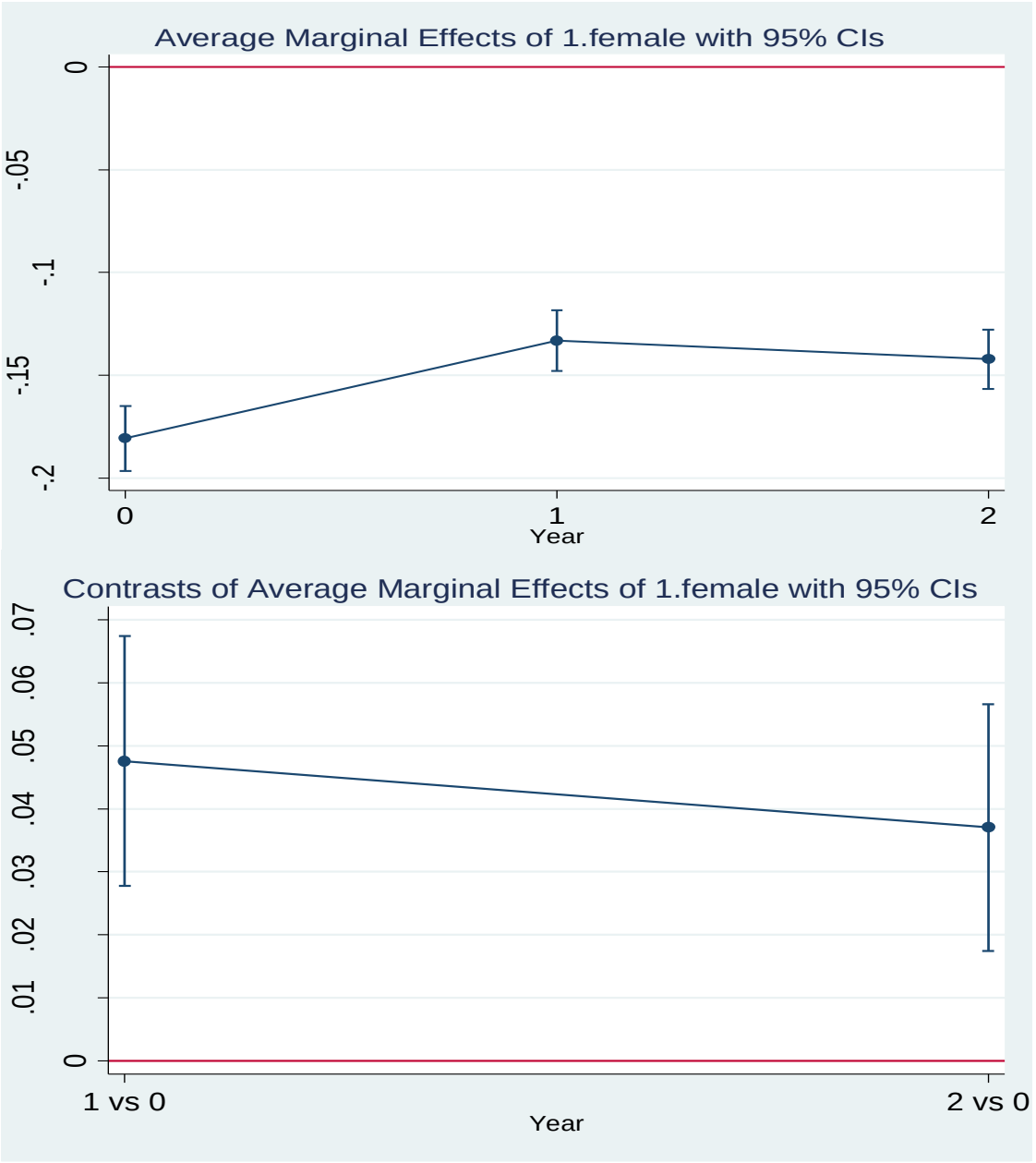


Estimates of the average (marginal) effect of being female on standardised test scores

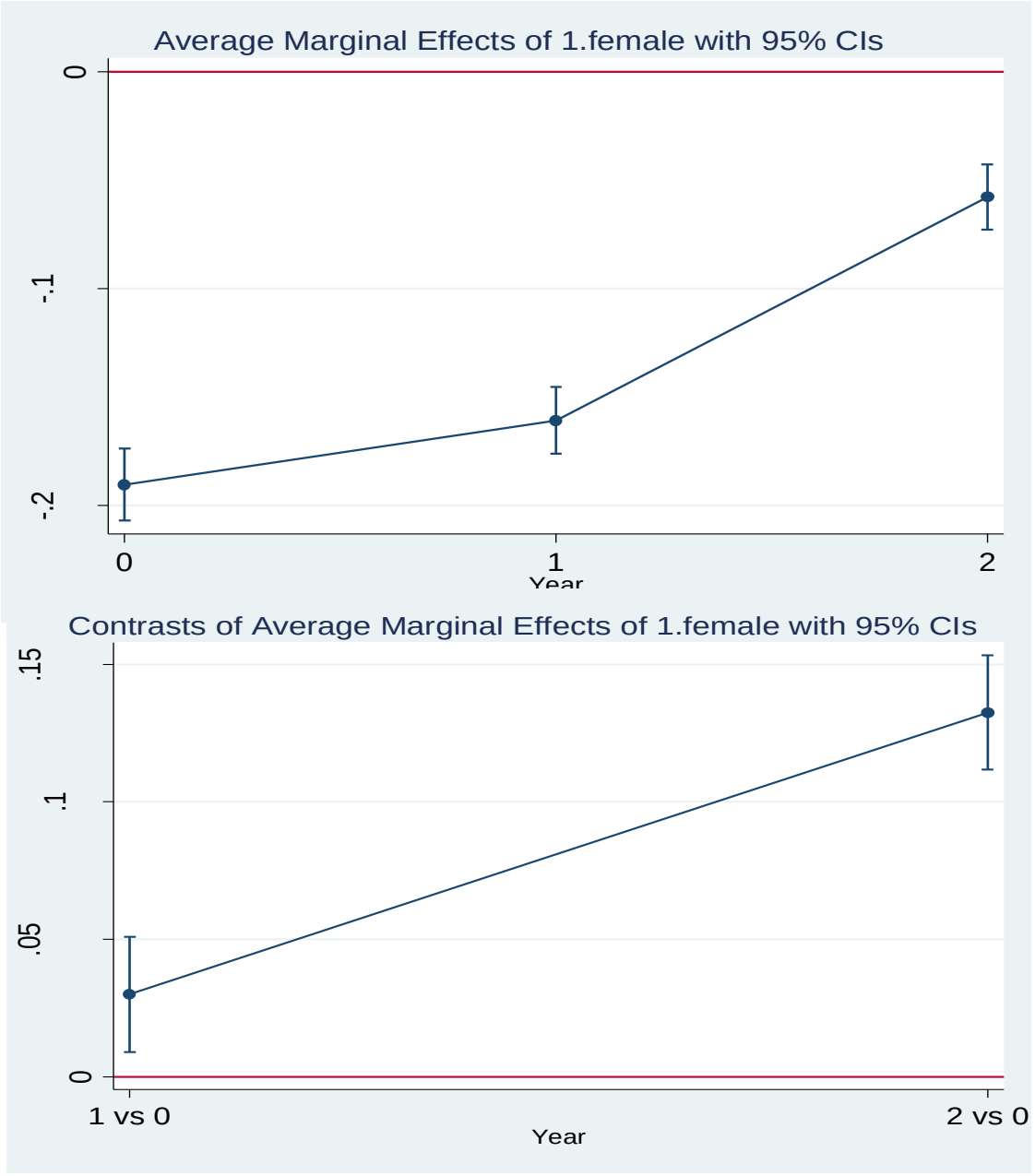


Comparison of the estimates of the average (marginal) effect of being female on standardised test scores across years (2011=0)

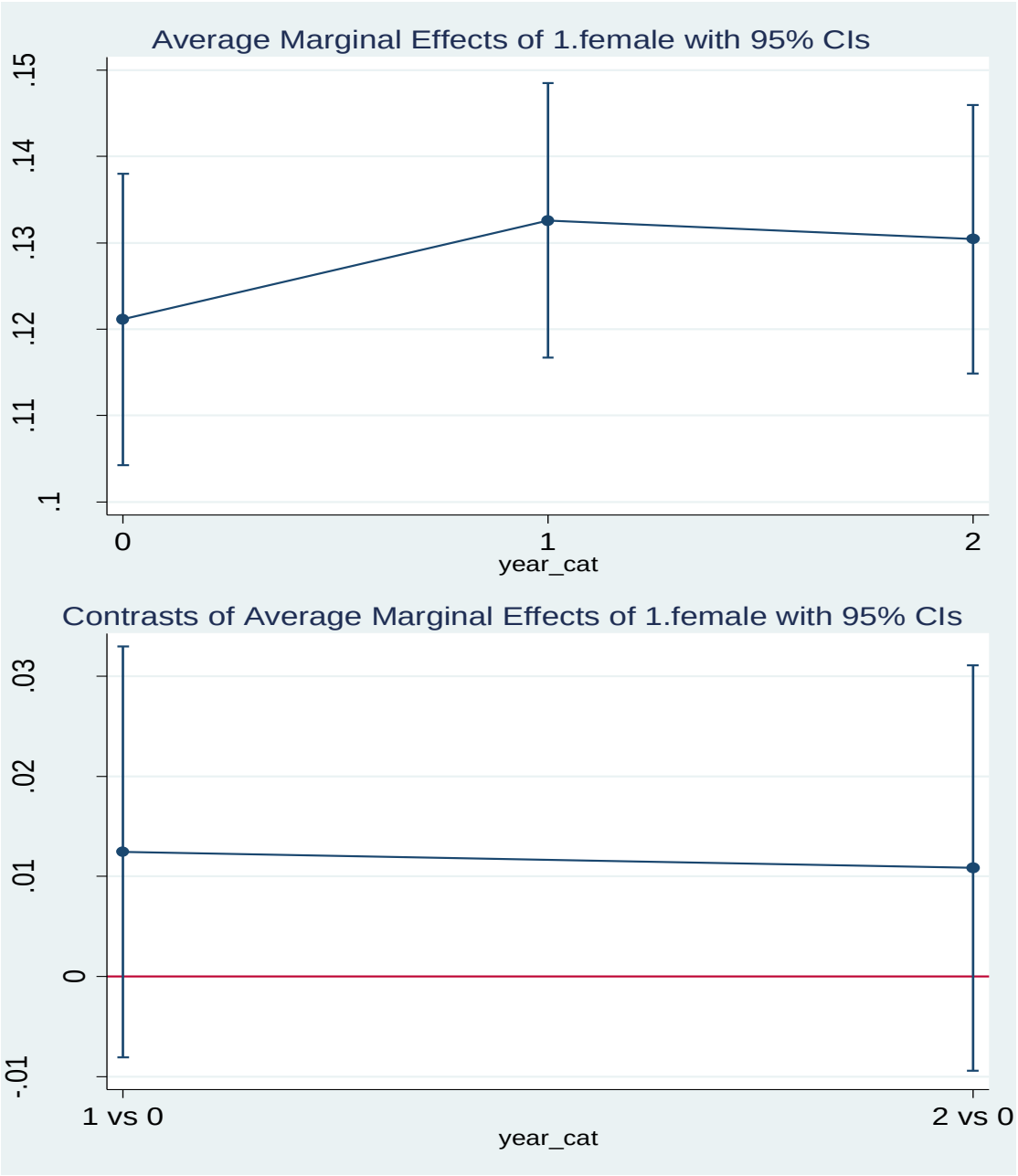
Overall female effect - Maths



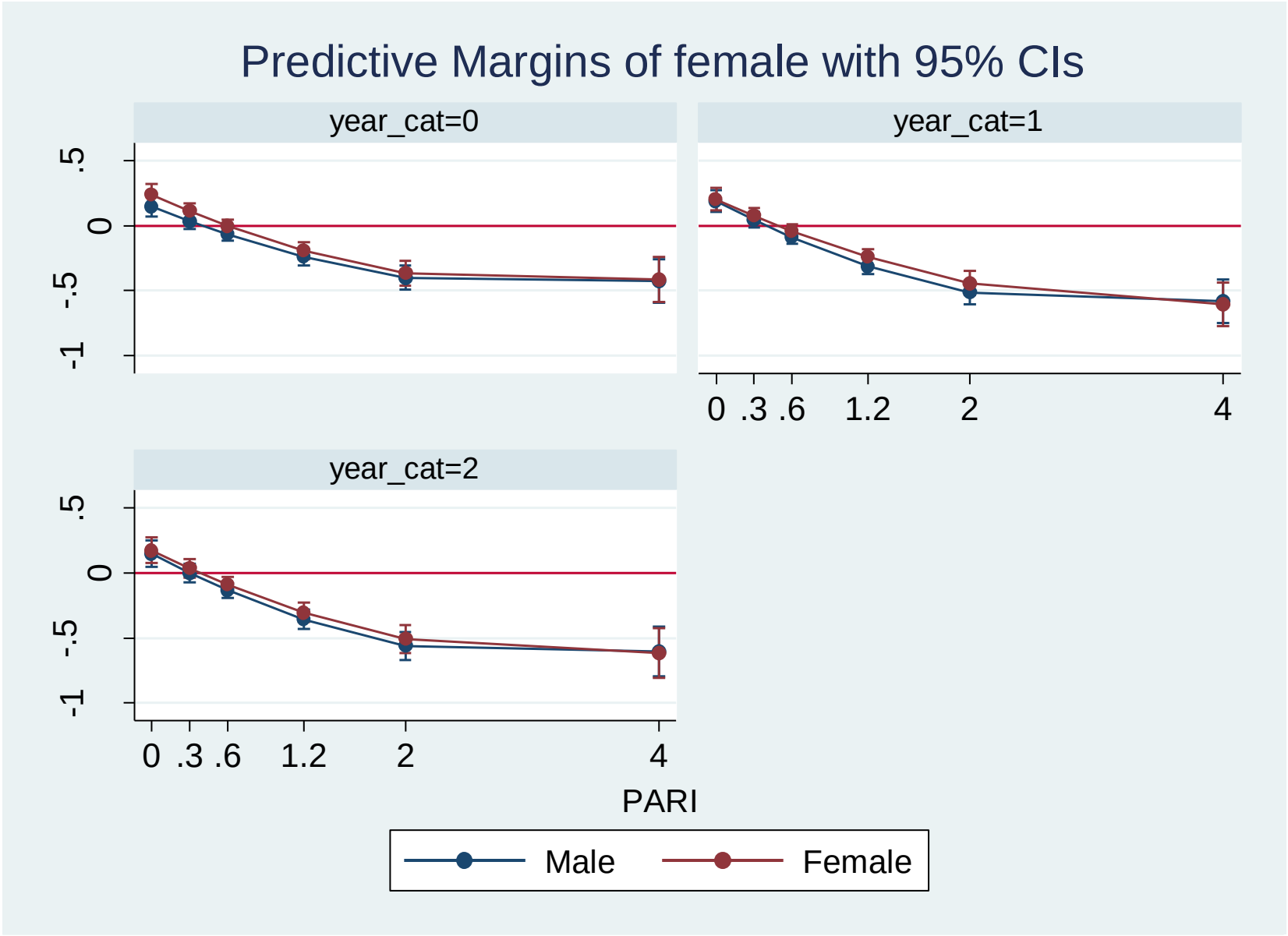
Overall female effect - Combined



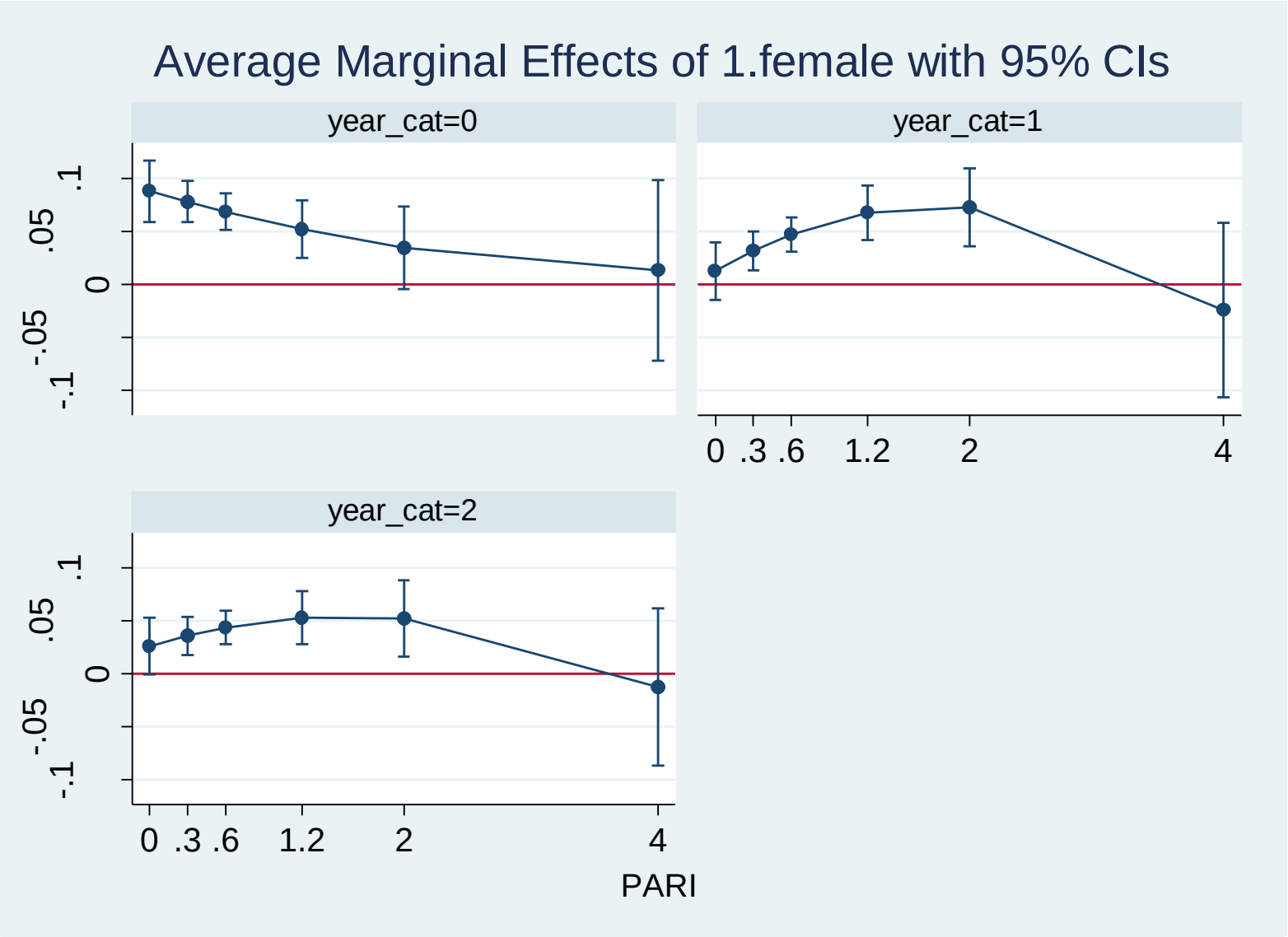
Overall female effect - ELP



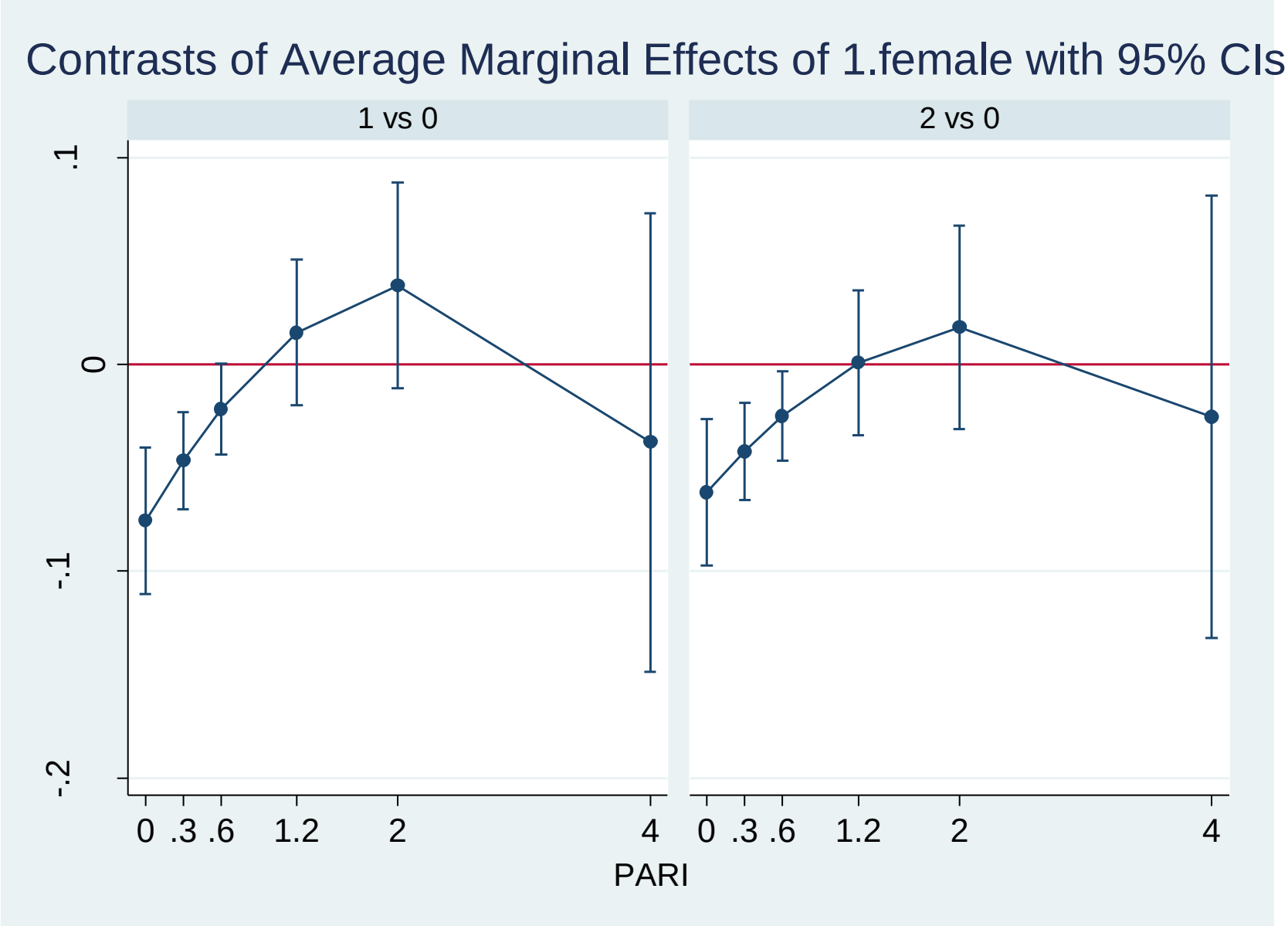
Female effect cond. remoteness - English



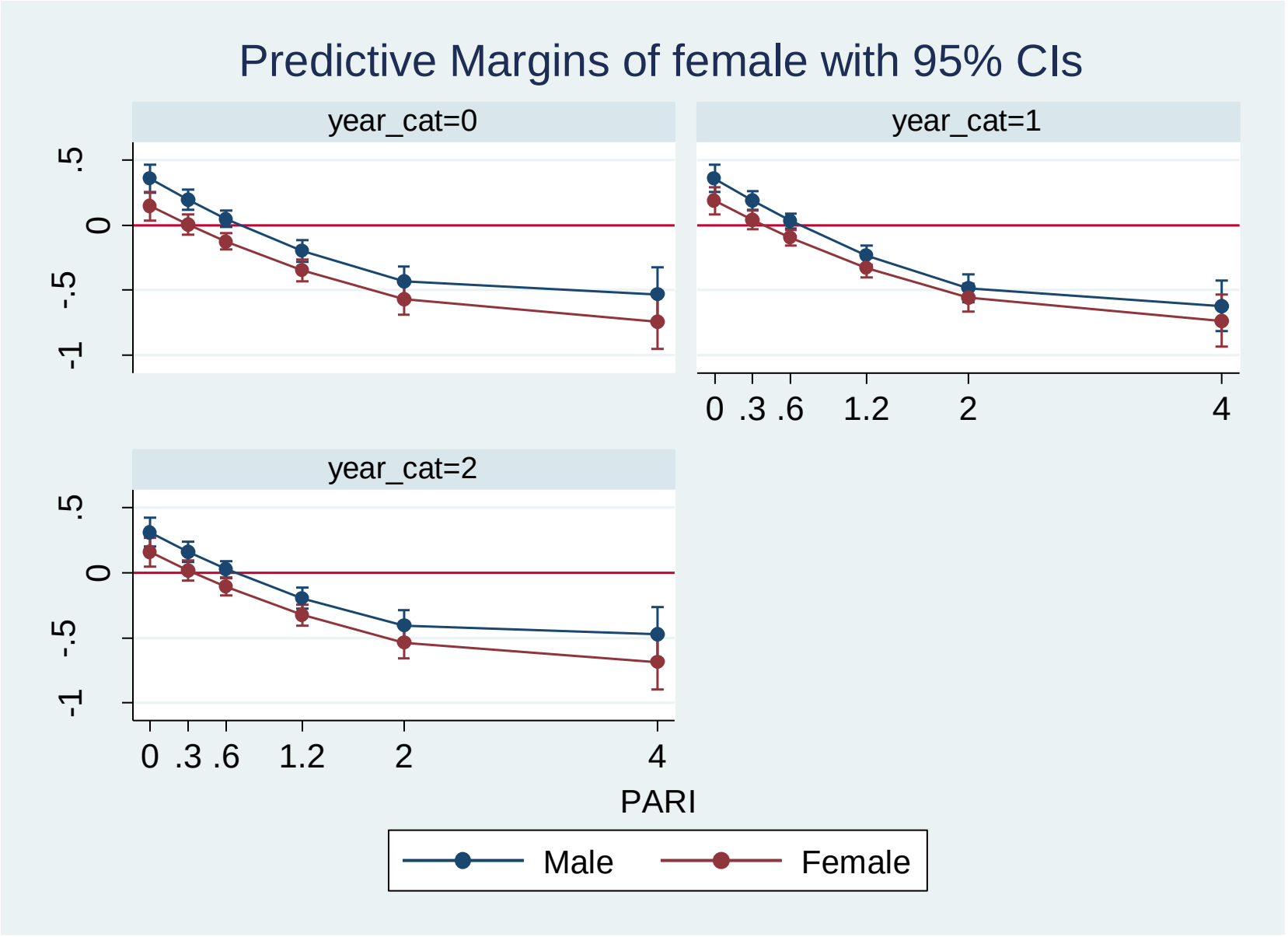
Female effect cond. remoteness - English



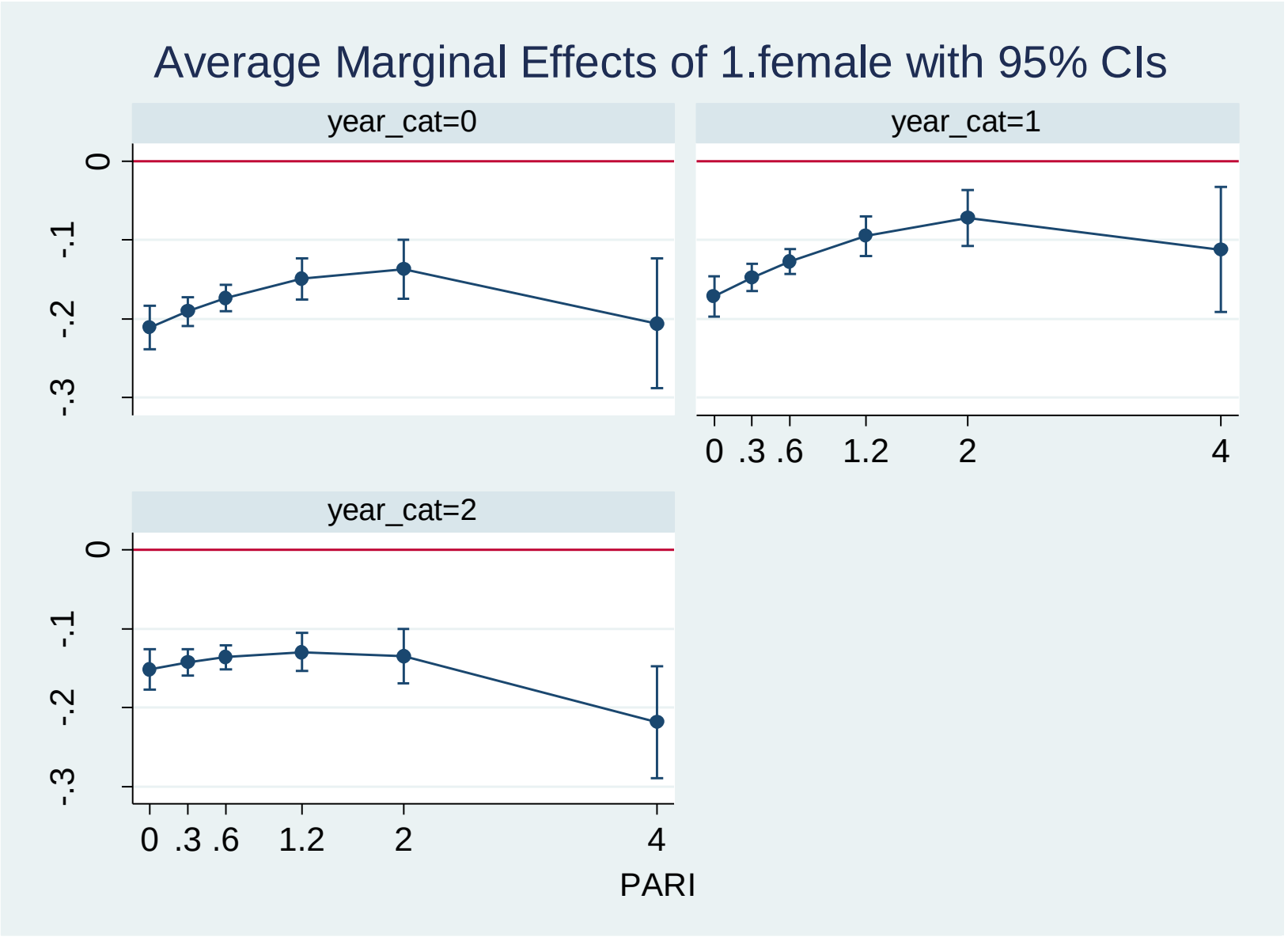
Female effect cond. remoteness - English



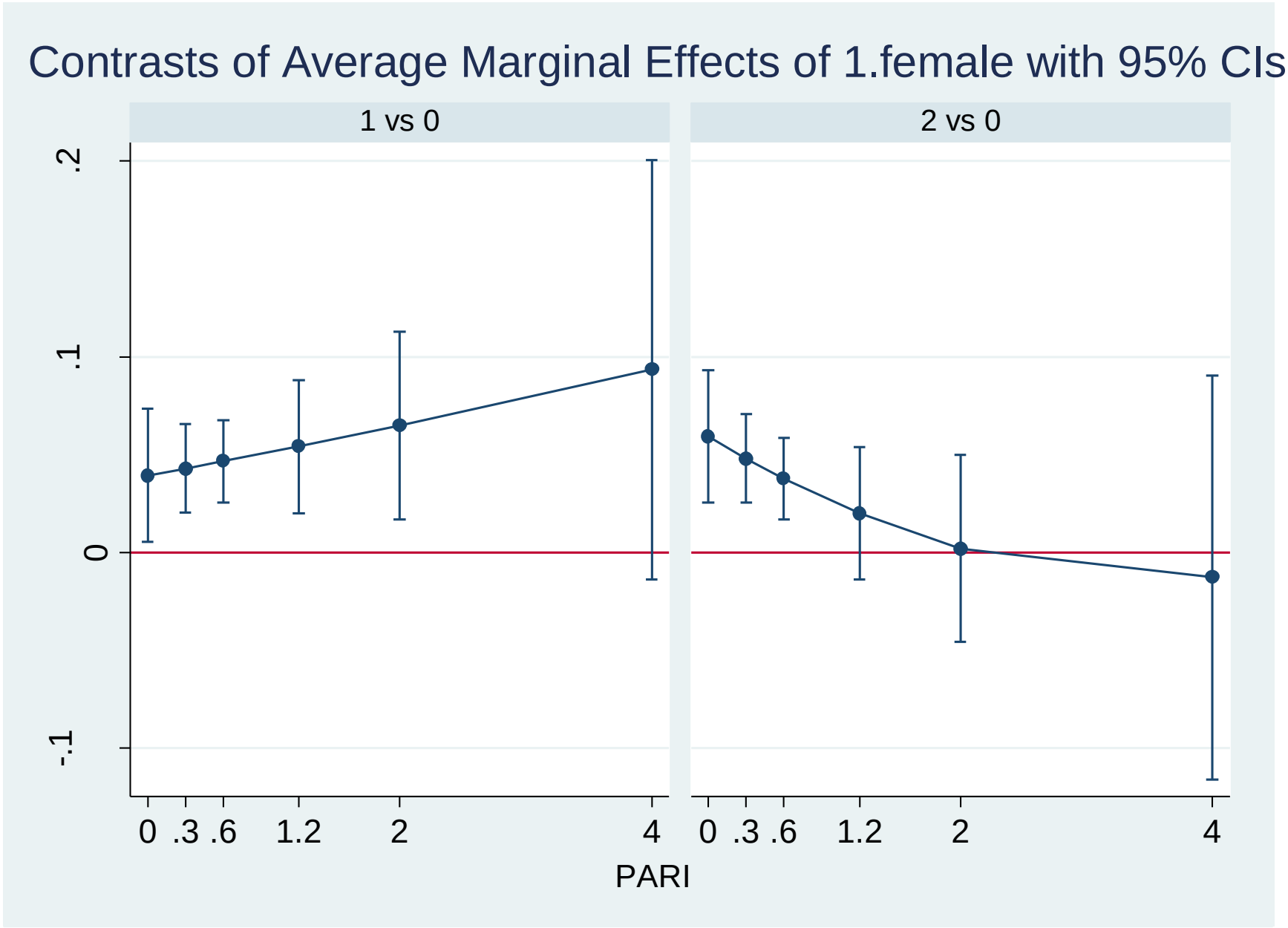
Female effect cond. remoteness - Maths



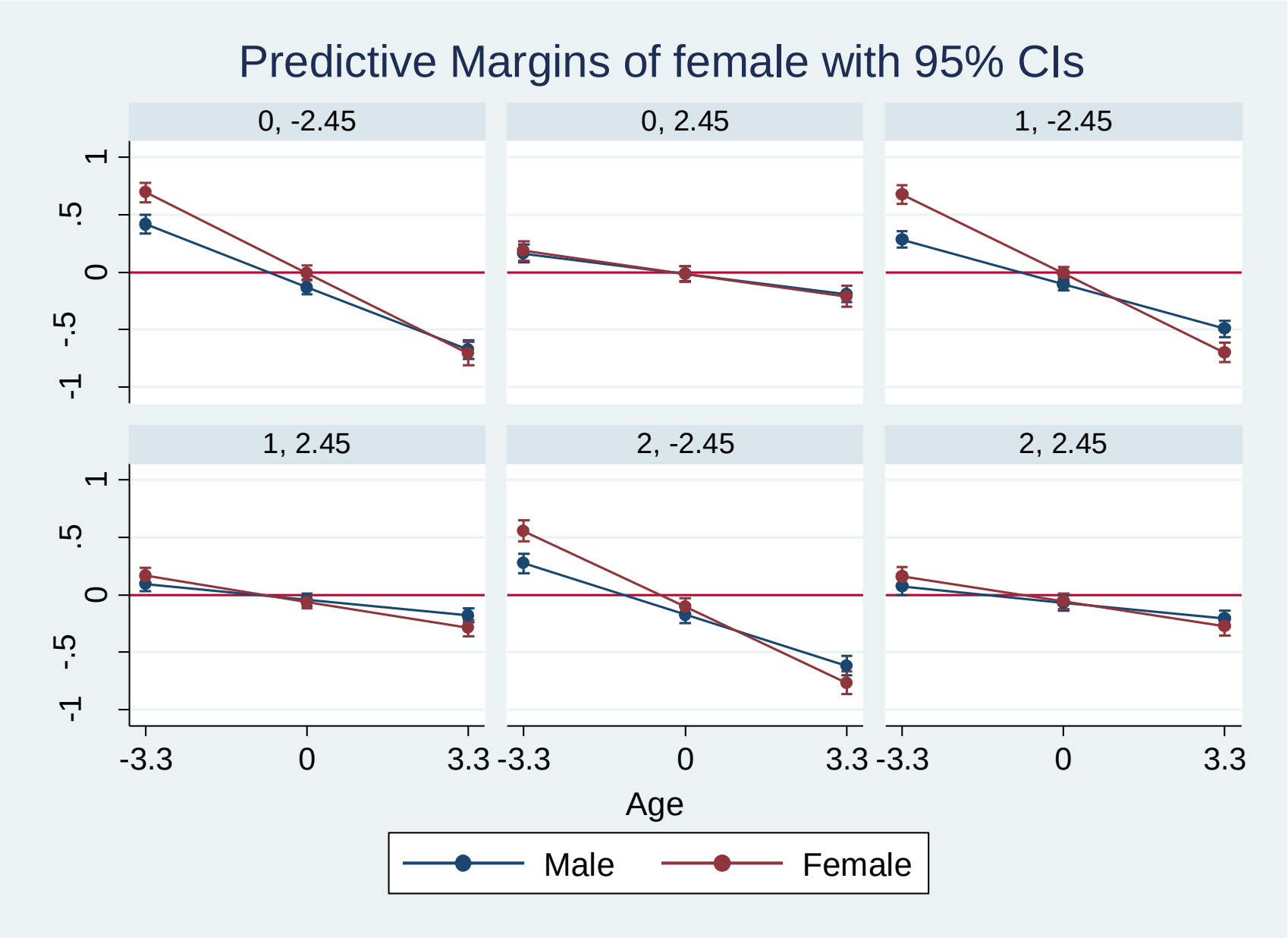
Female effect cond. remoteness - Maths



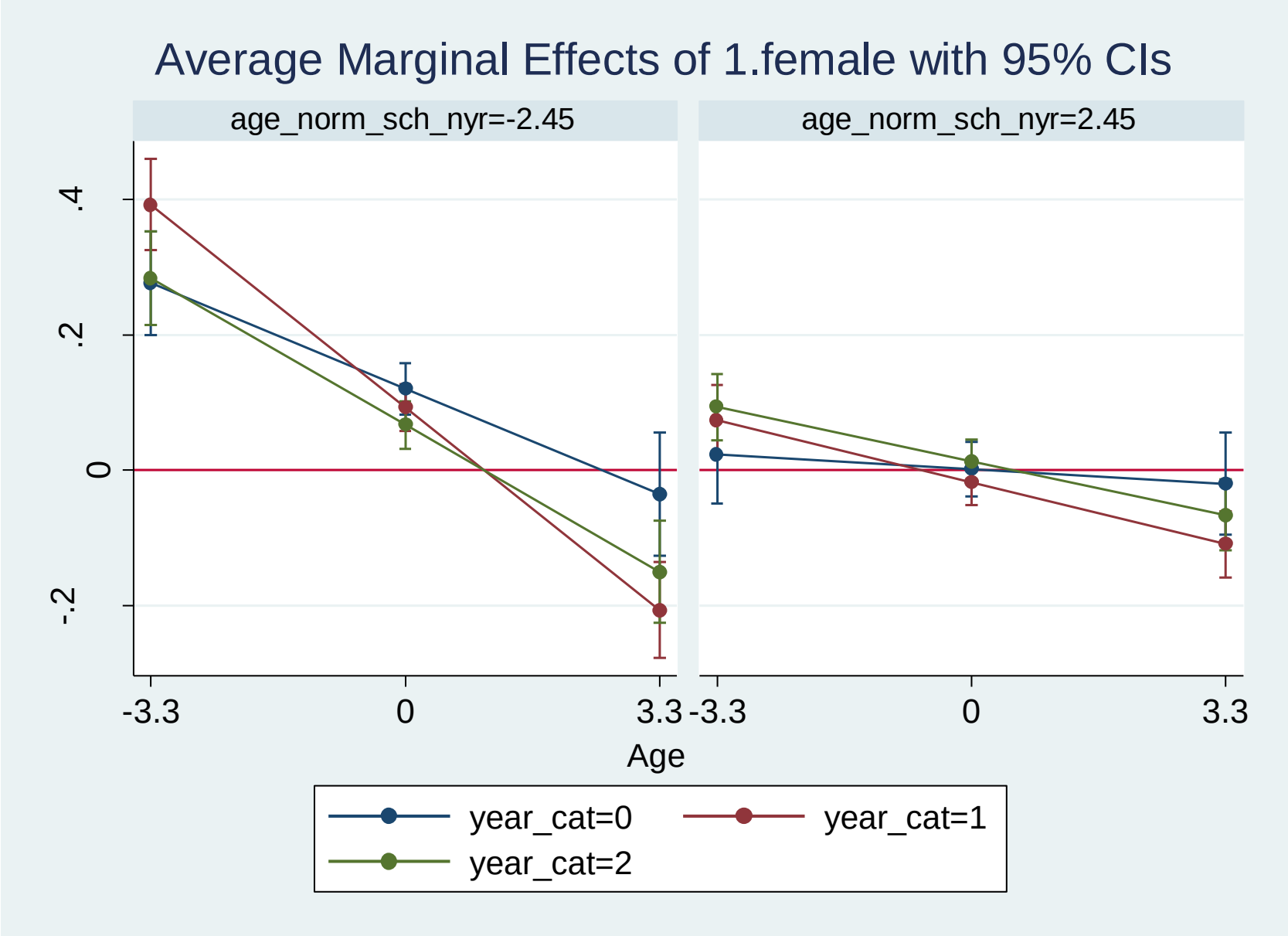
Female effect cond. remoteness - Maths



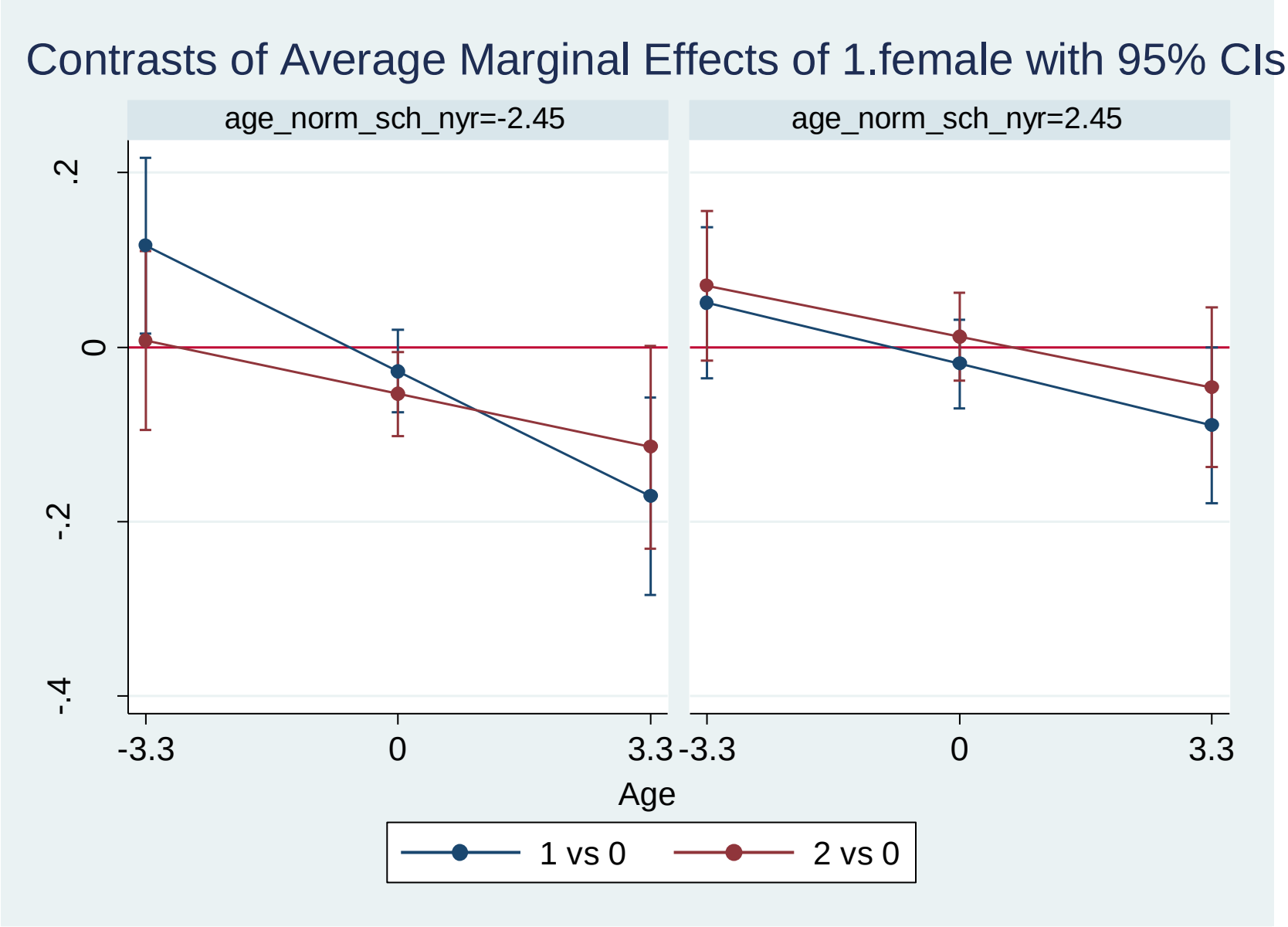
Female effect: age, av. school age- English



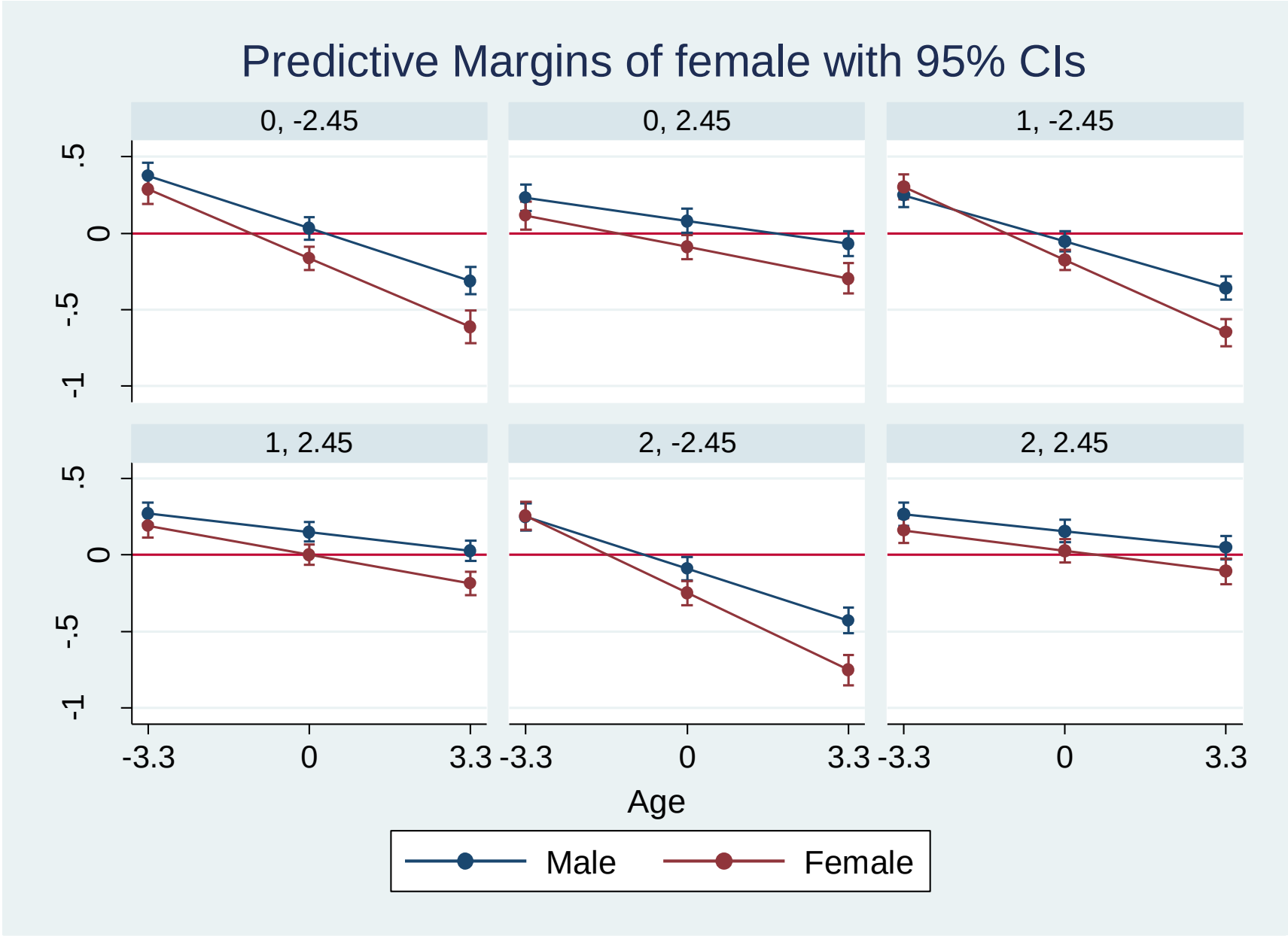
Female effect: age, av. school age- English



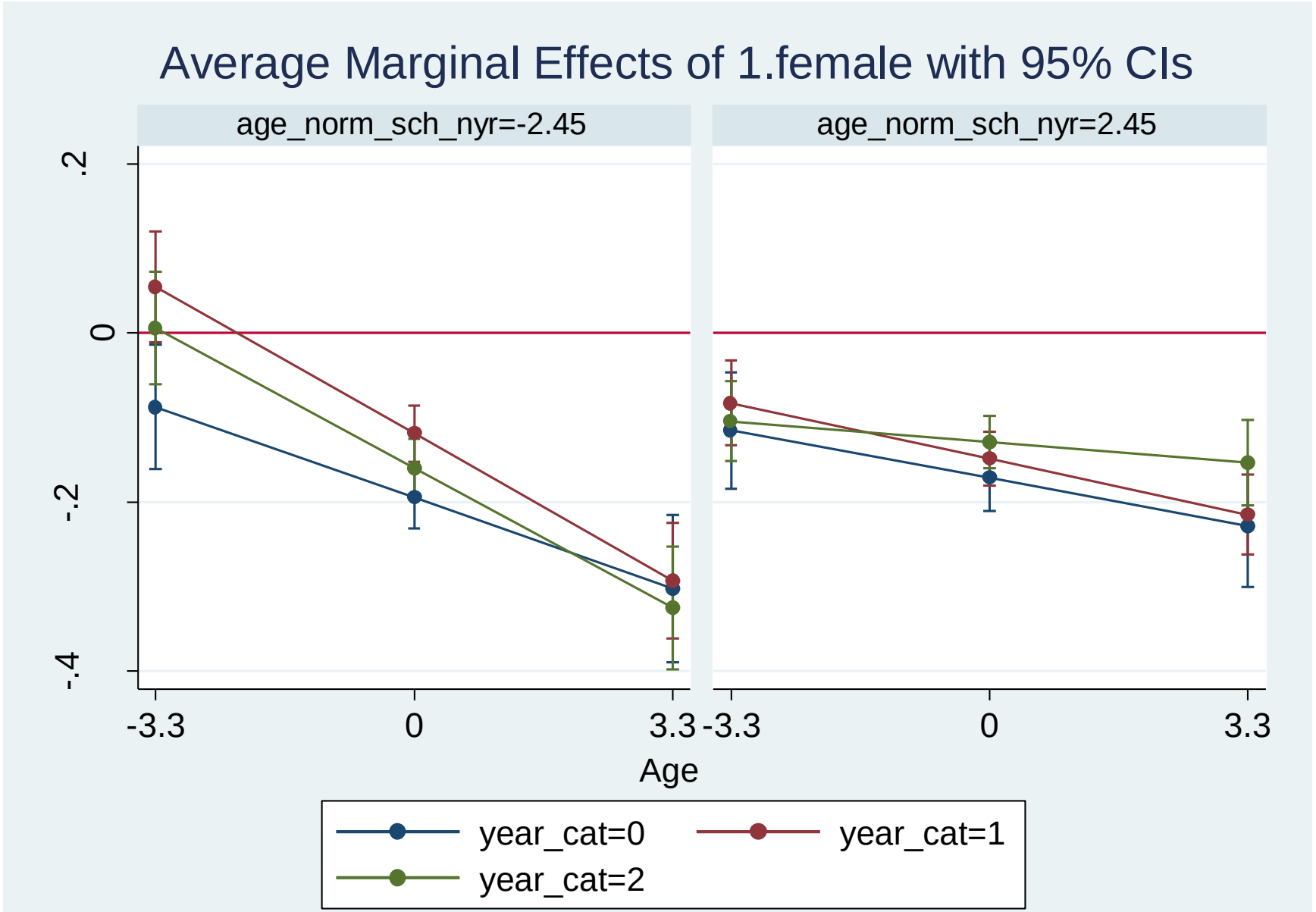
Female effect: age, av. school age- English



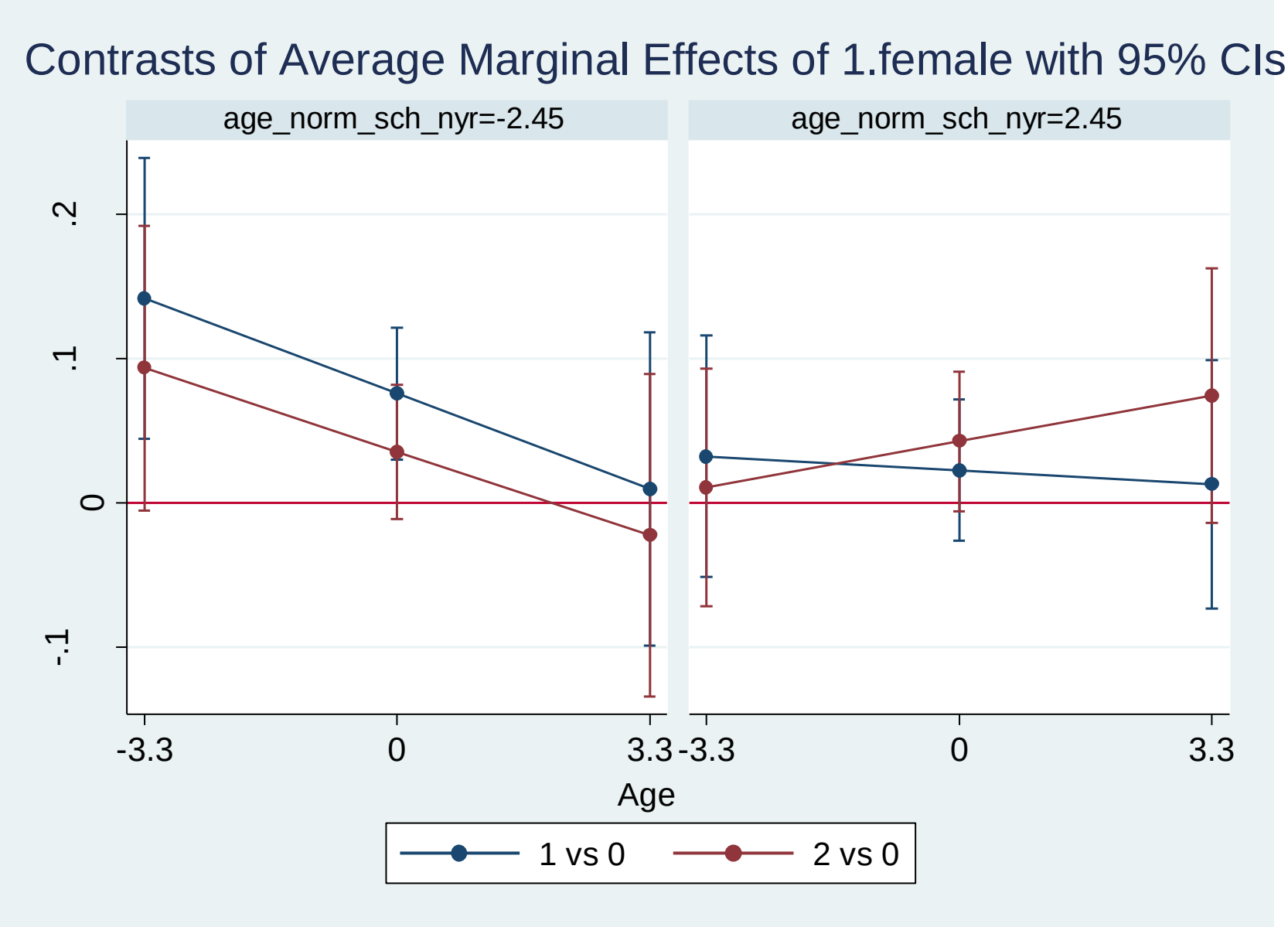
Female effect: age, av. school age- Maths



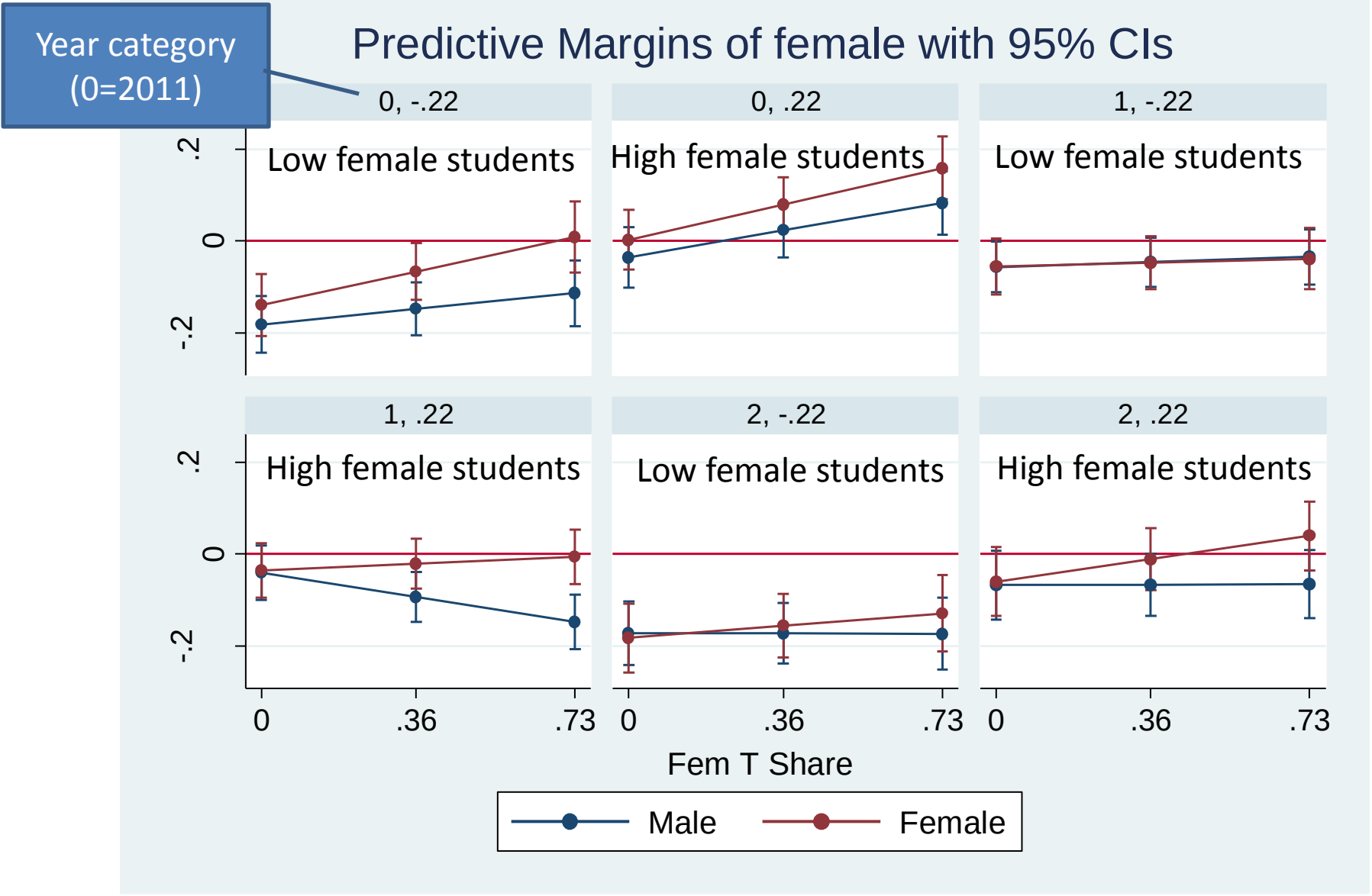
Female effect: age, av. school age- Maths



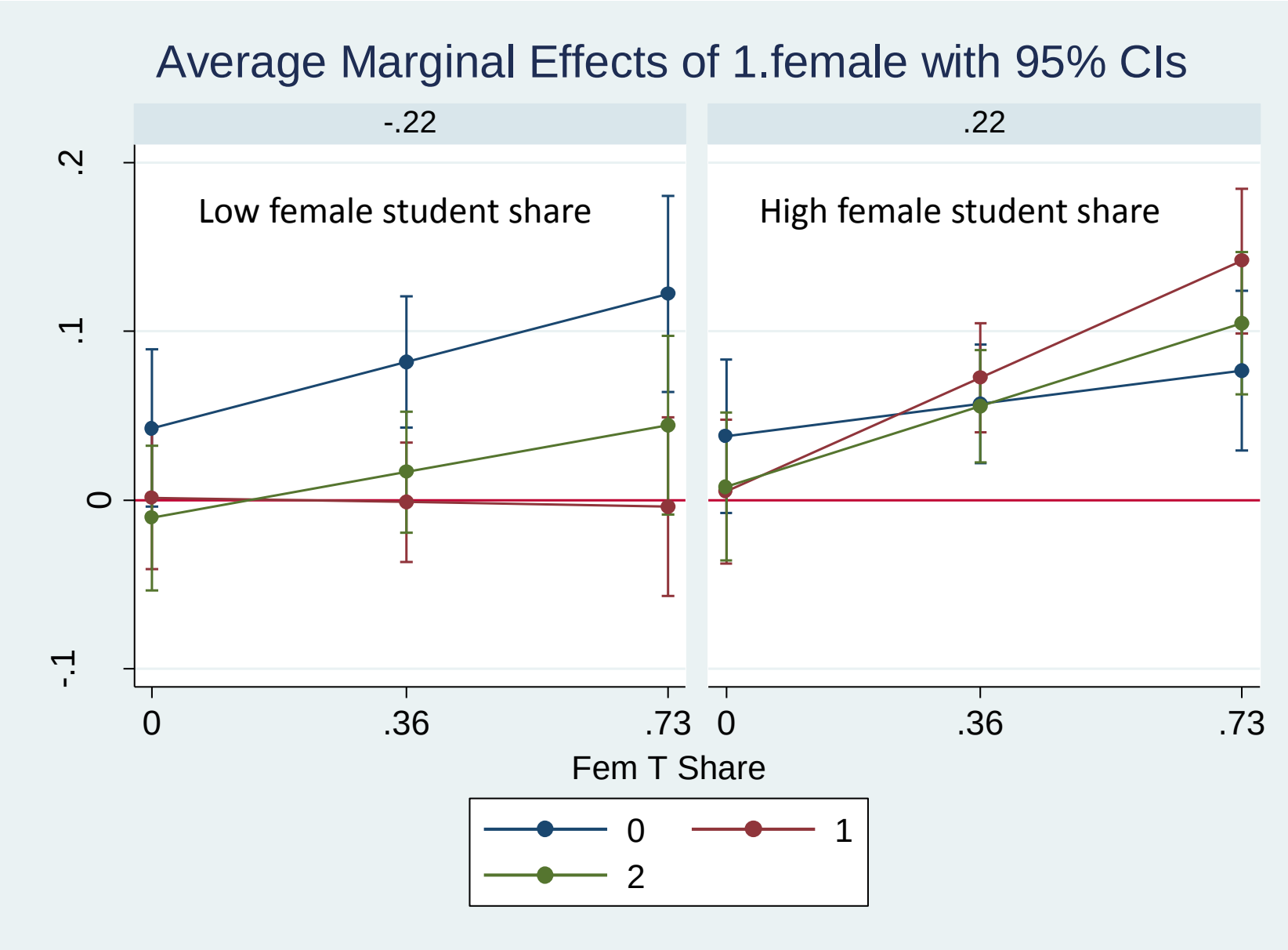
Female effect: age, av. school age- Maths



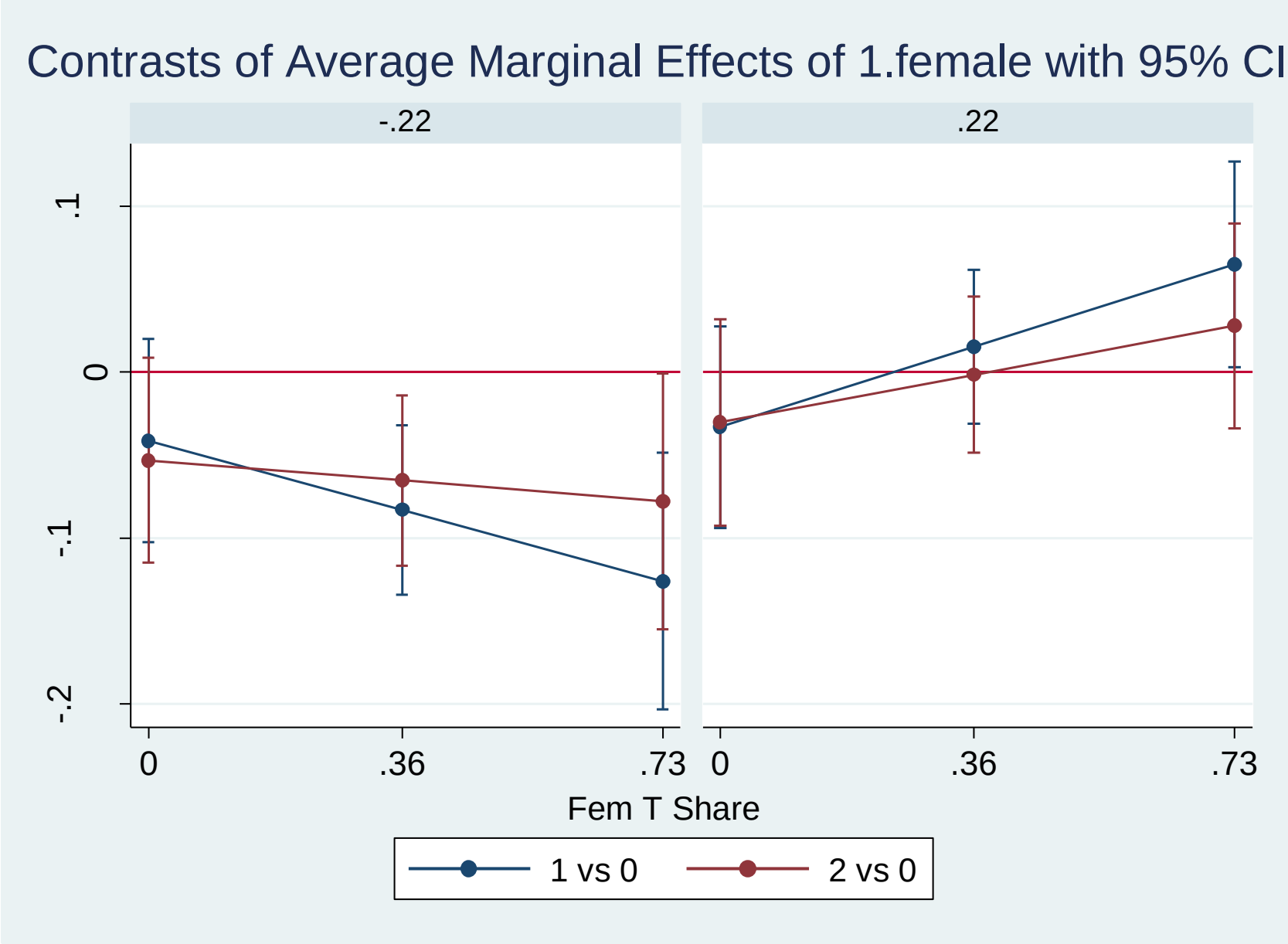
Female effect: female T & Students - English



Female effect: female T & Students - English

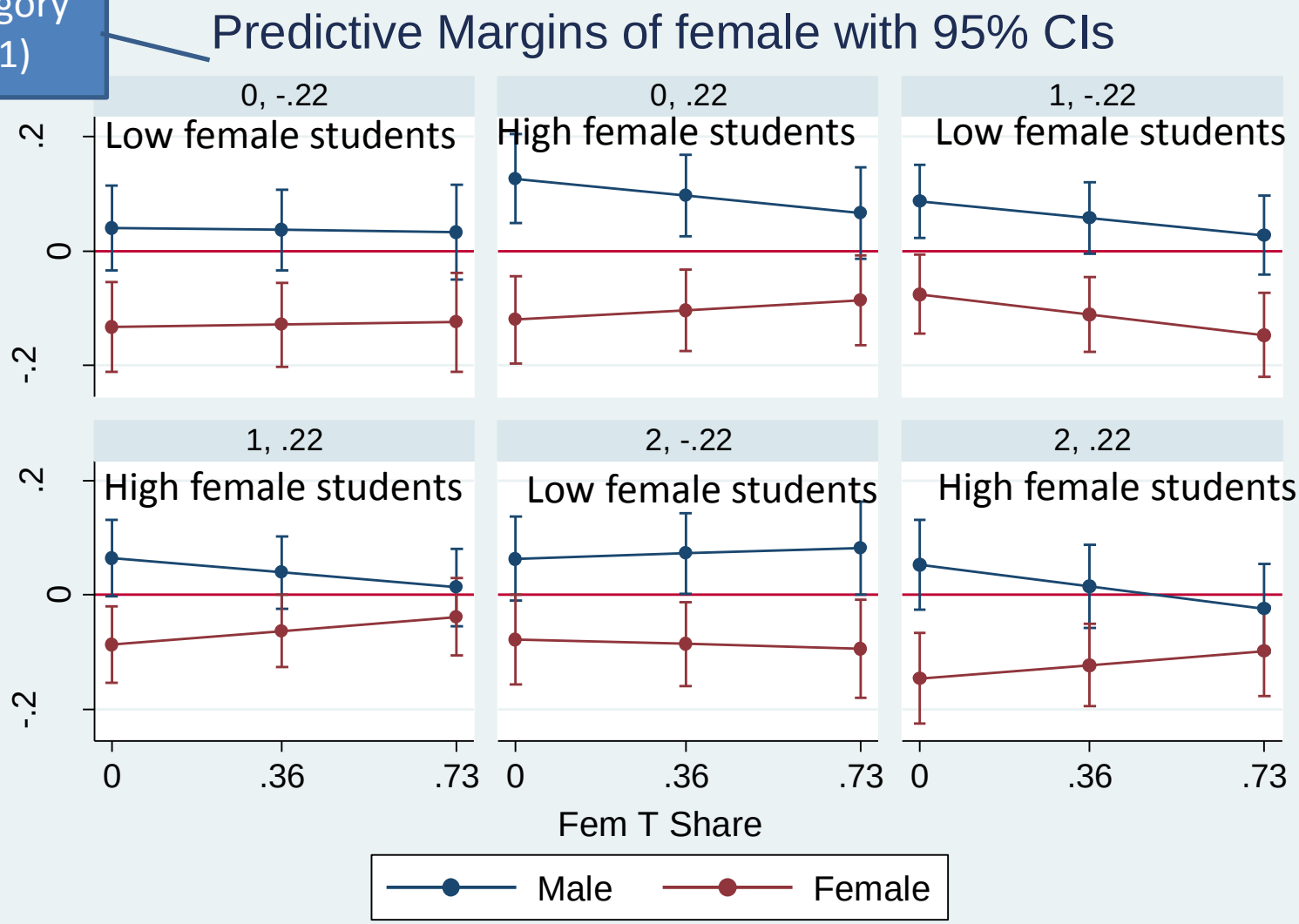


Female effect: female T & Students - English

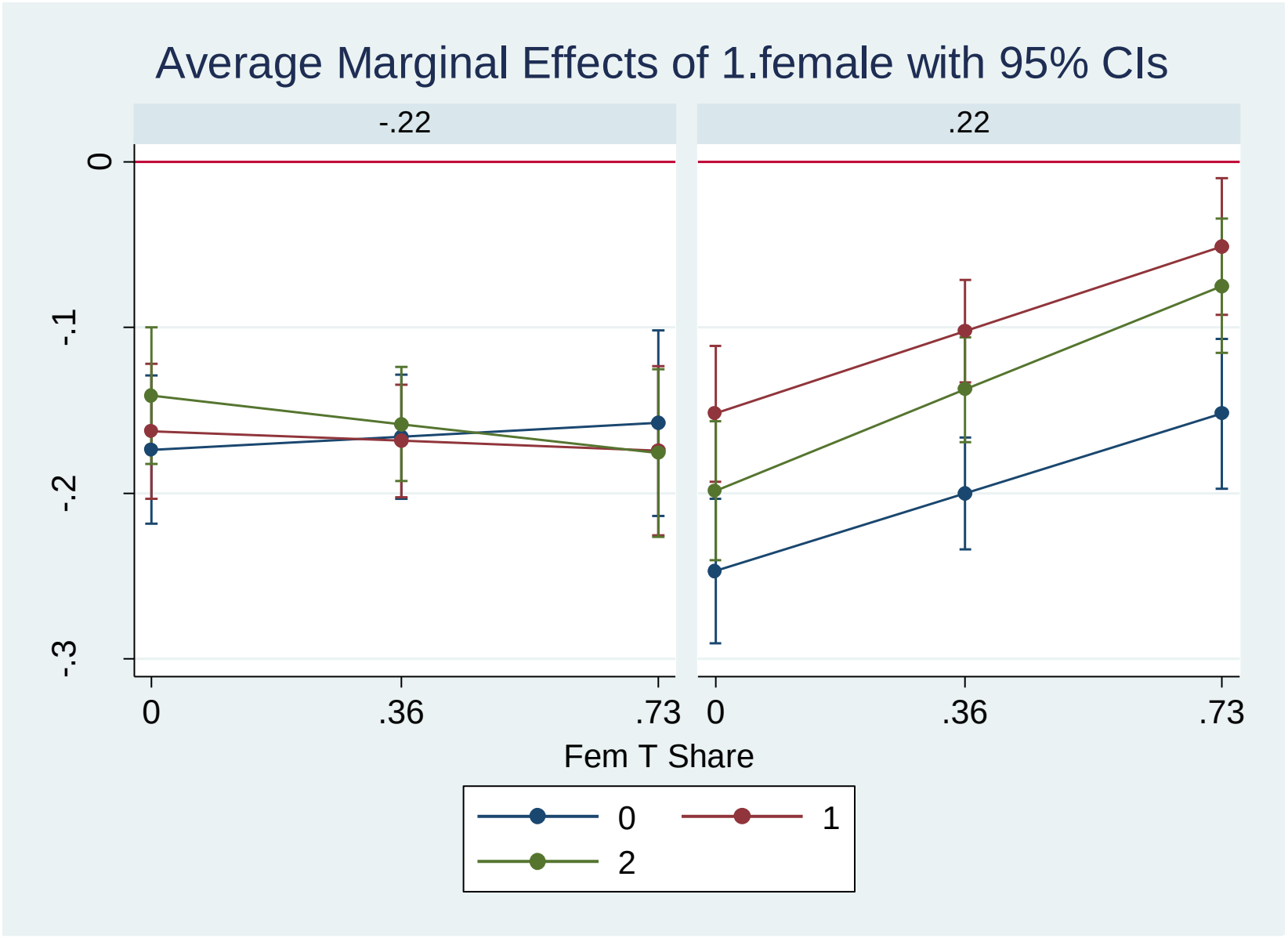


Female effect: female T & Students - Maths

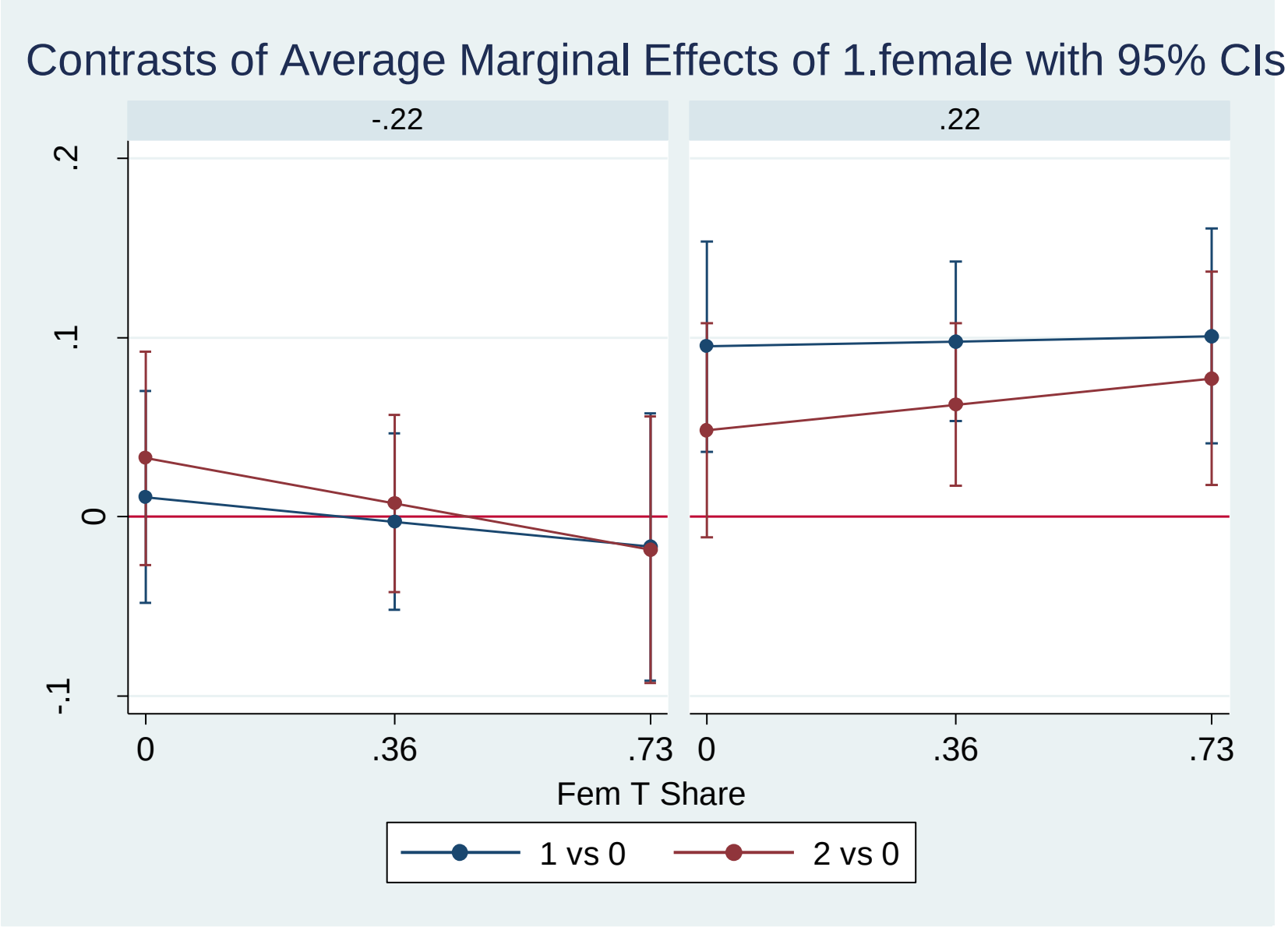
Year category (0=2011)



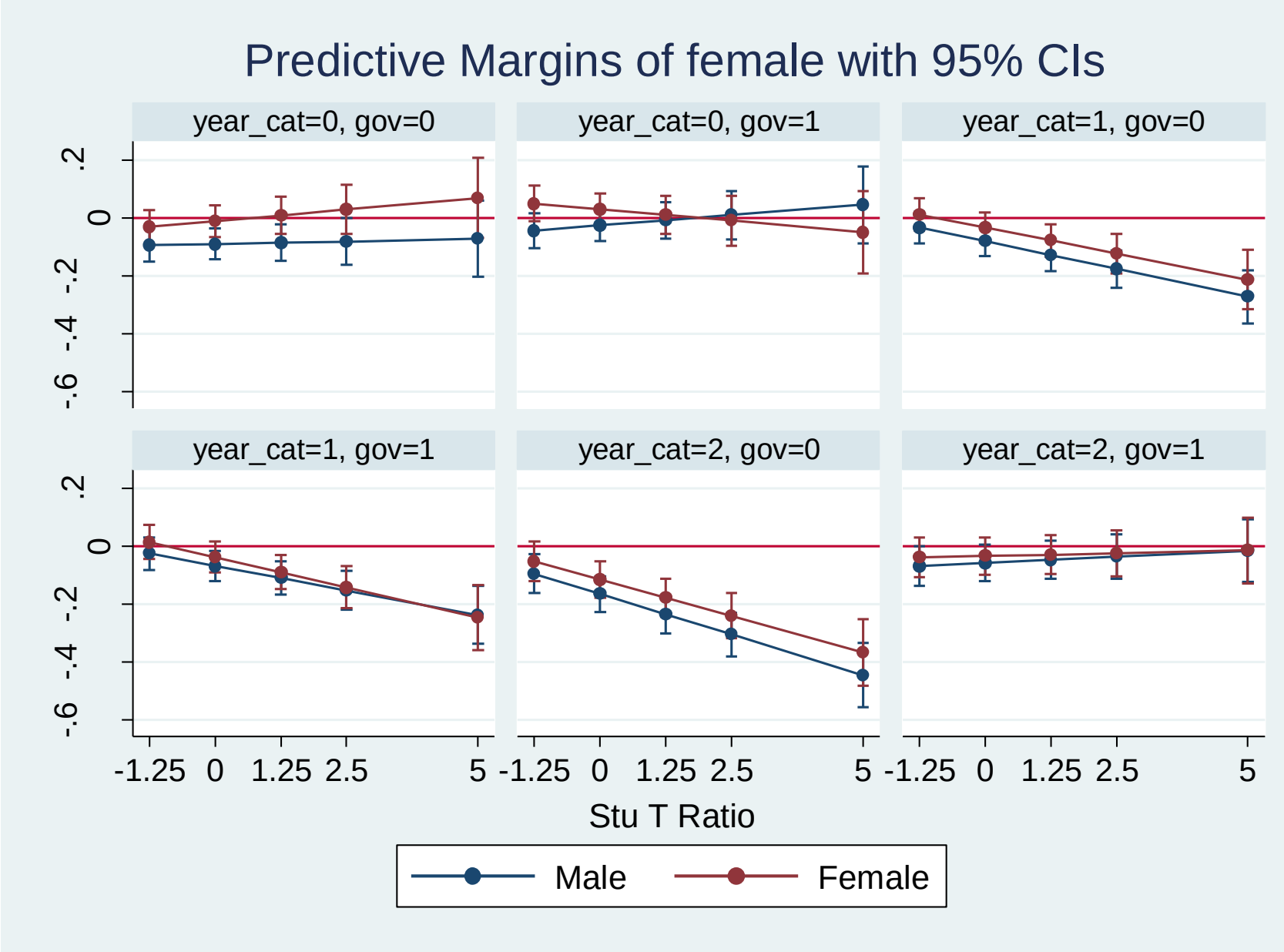
Female effect: female T & Students - Maths



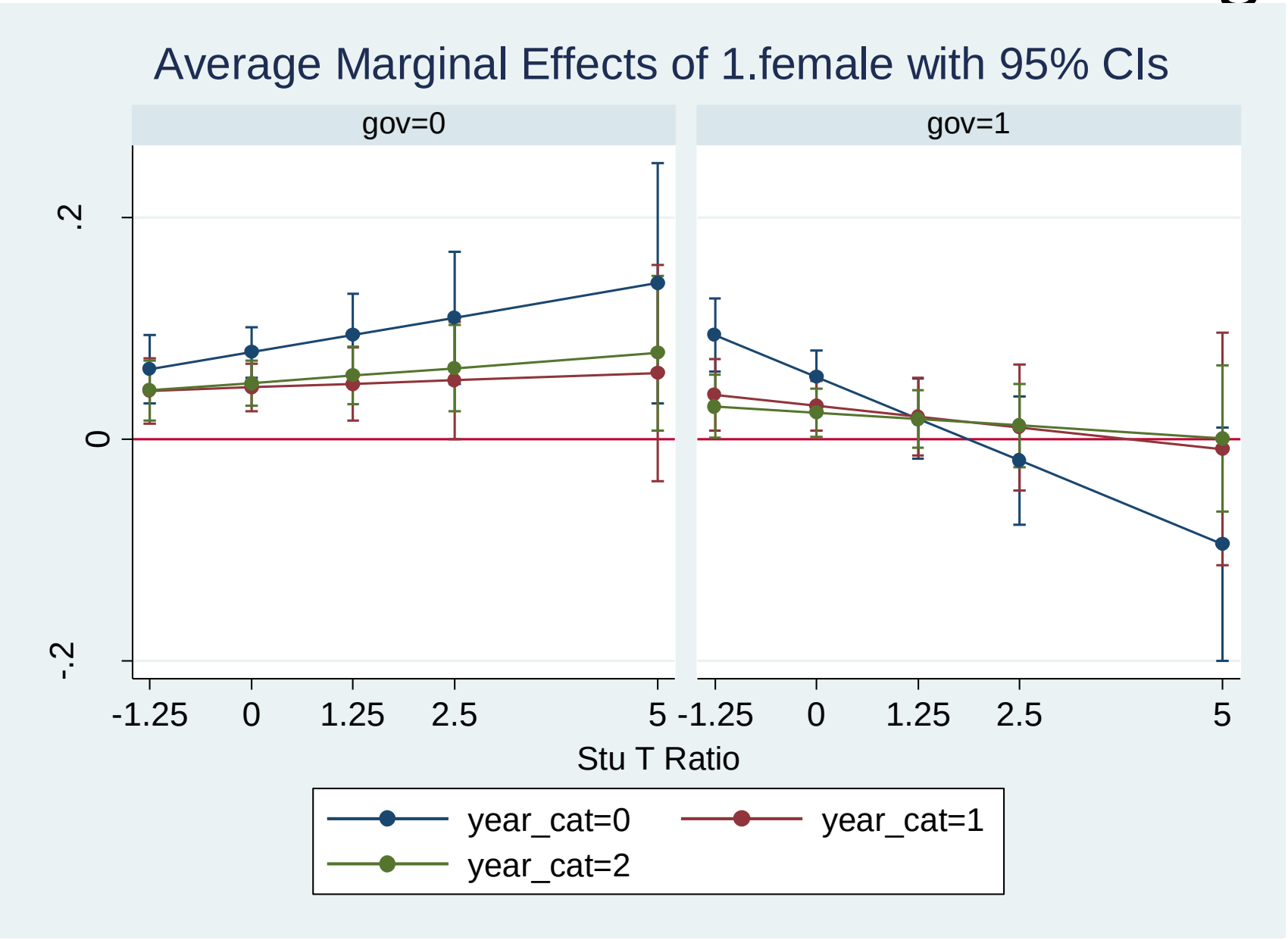
Female effect: female T & Students - Maths



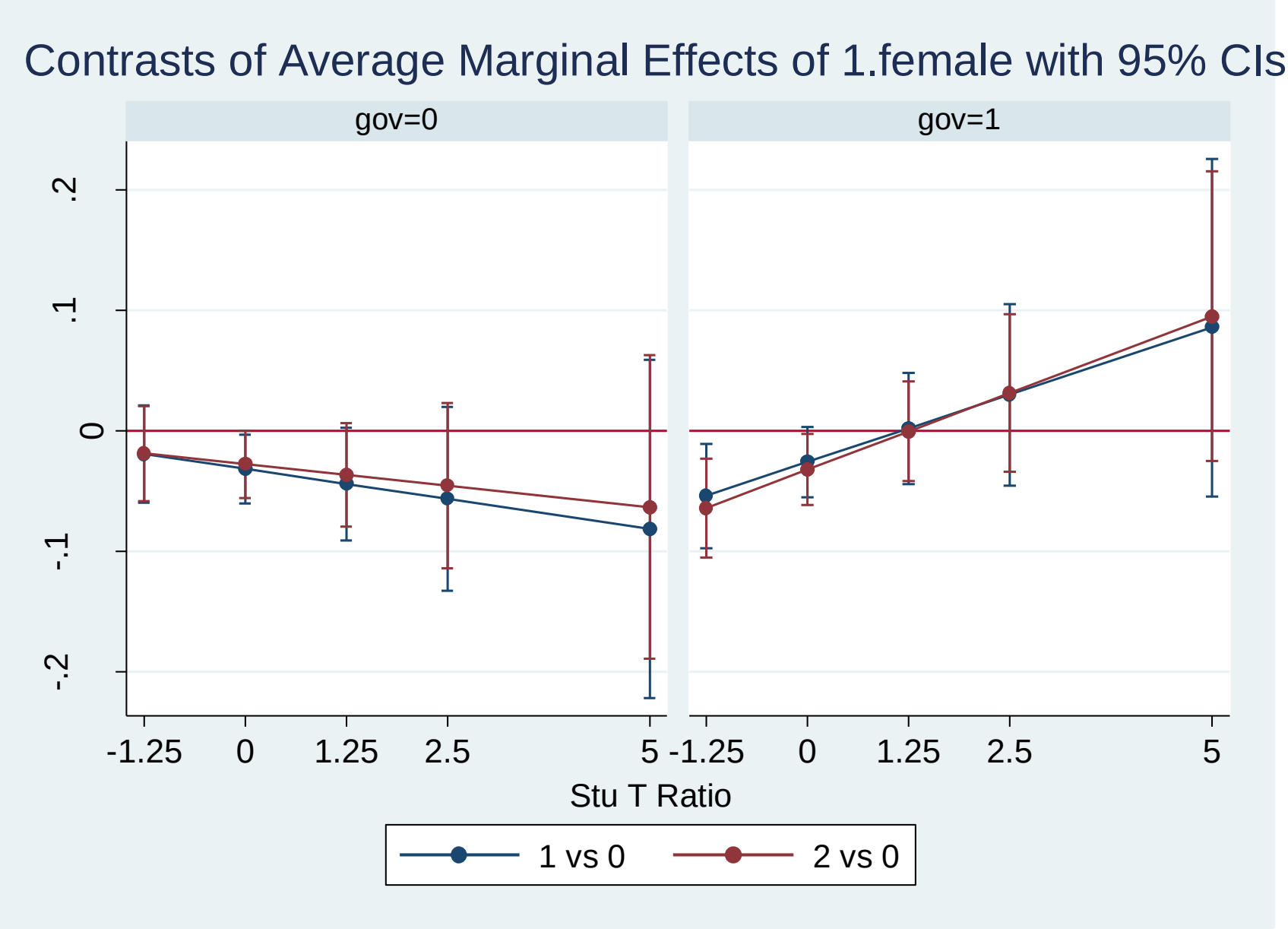
Female effect: Class size & Govt- English



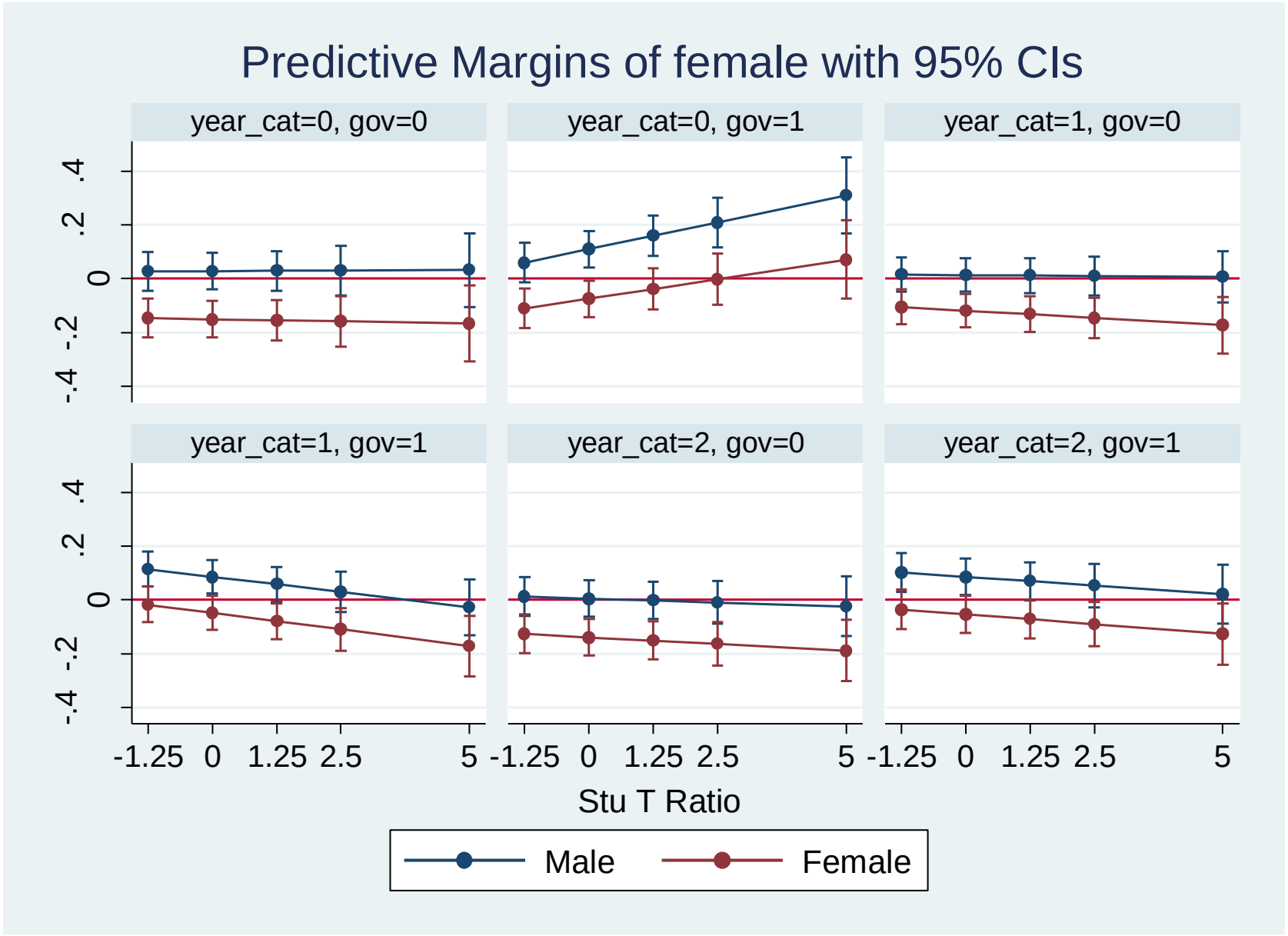
Female effect: Class size & Govt- English



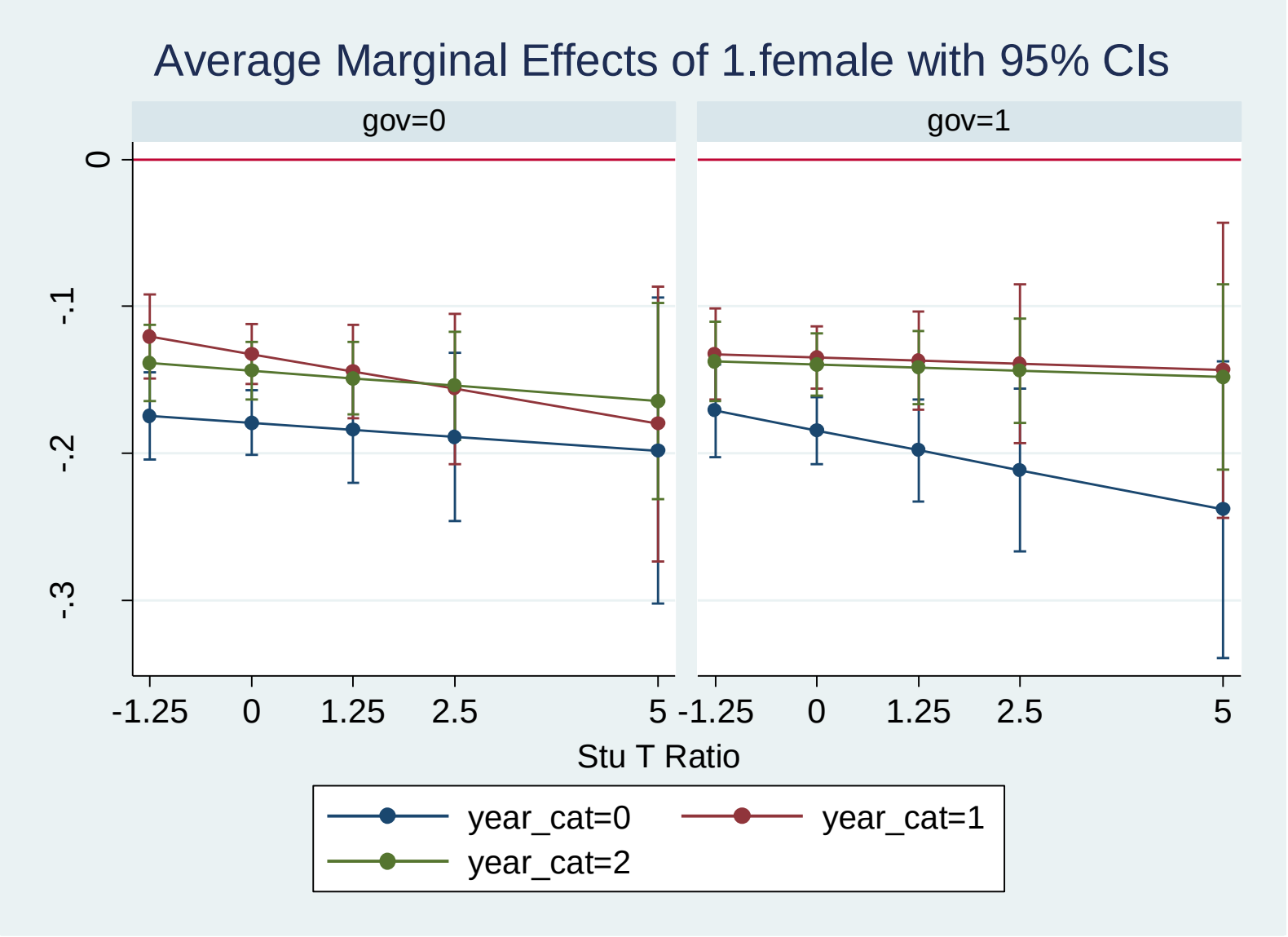
Female effect: Class size & Govt- English



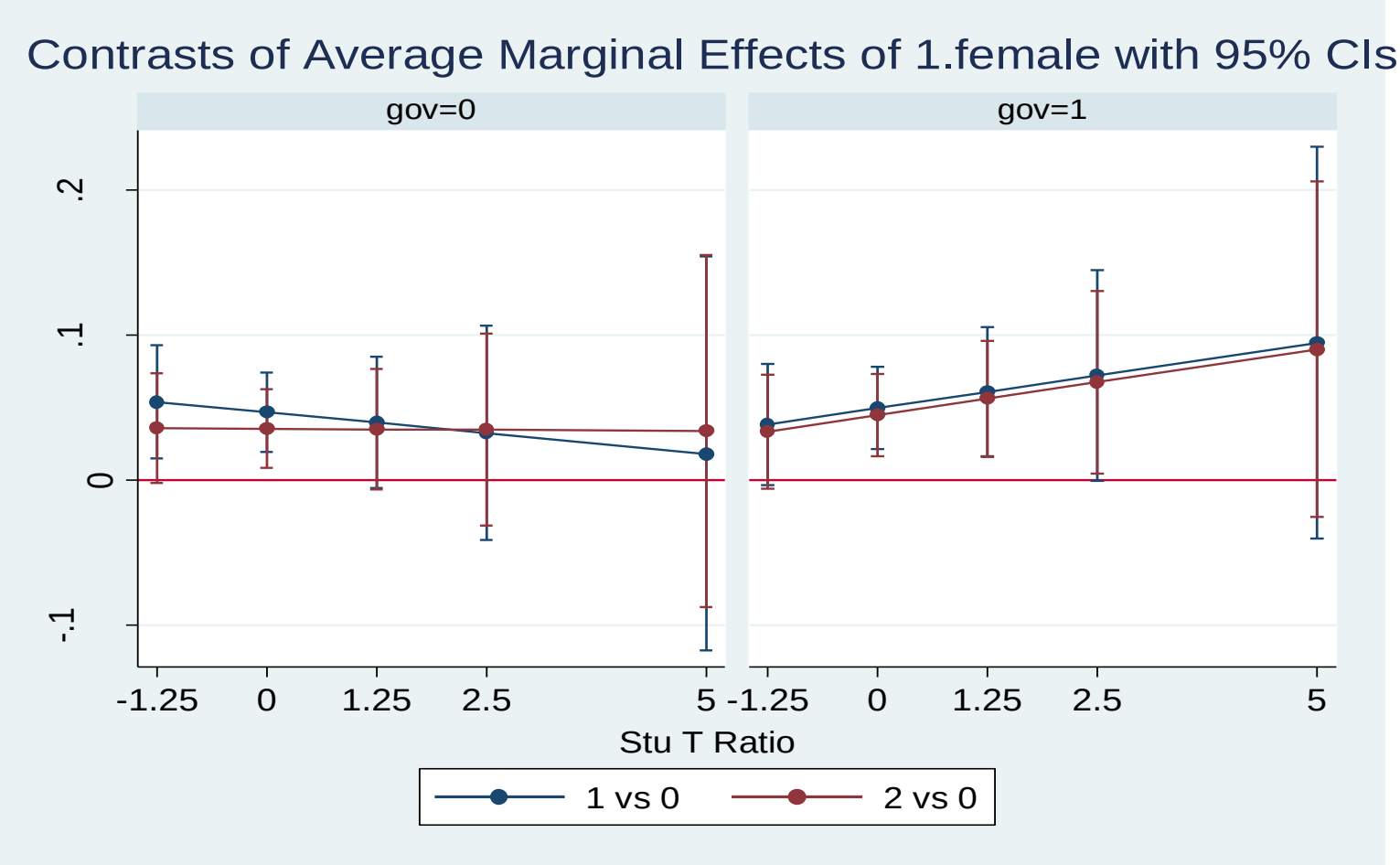
Female effect: Class size & Govt- Maths



Female effect: Class size & Govt- Maths



Female effect: Class size & Govt- Maths



An improvement in gender outcomes in average sized classes in both church and government schools.

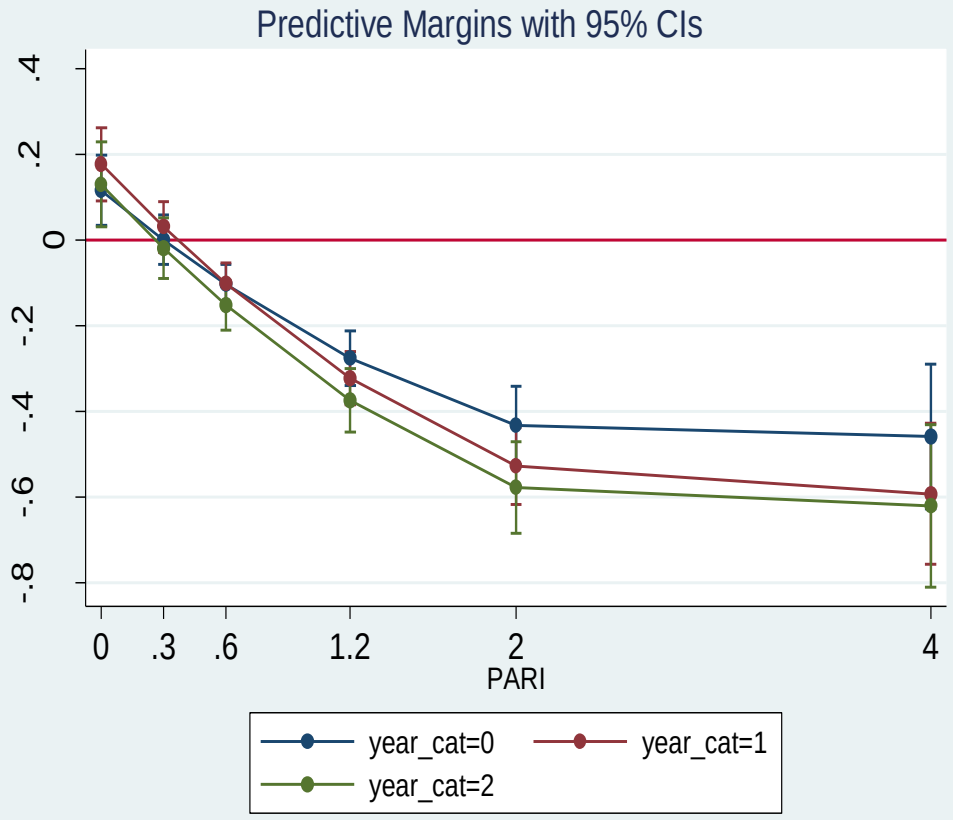
What might explain gender effects?

- Female effect is increasing with remoteness, female teacher share, female student share; and decreasing with age, average class age, and slightly with class size.
- Female effect has increased after 2011 (with the introduction of the TFF) for Maths and Combined but decreased for English.
  - A decrease in the differences across gender.
- It's hard to explain effects opposite in sign using the same variables.
- Possible catch-up effect due to better resourced schools or better school based management outcomes.
- More students are coming into urban schools and females don't do as well in urban schools.

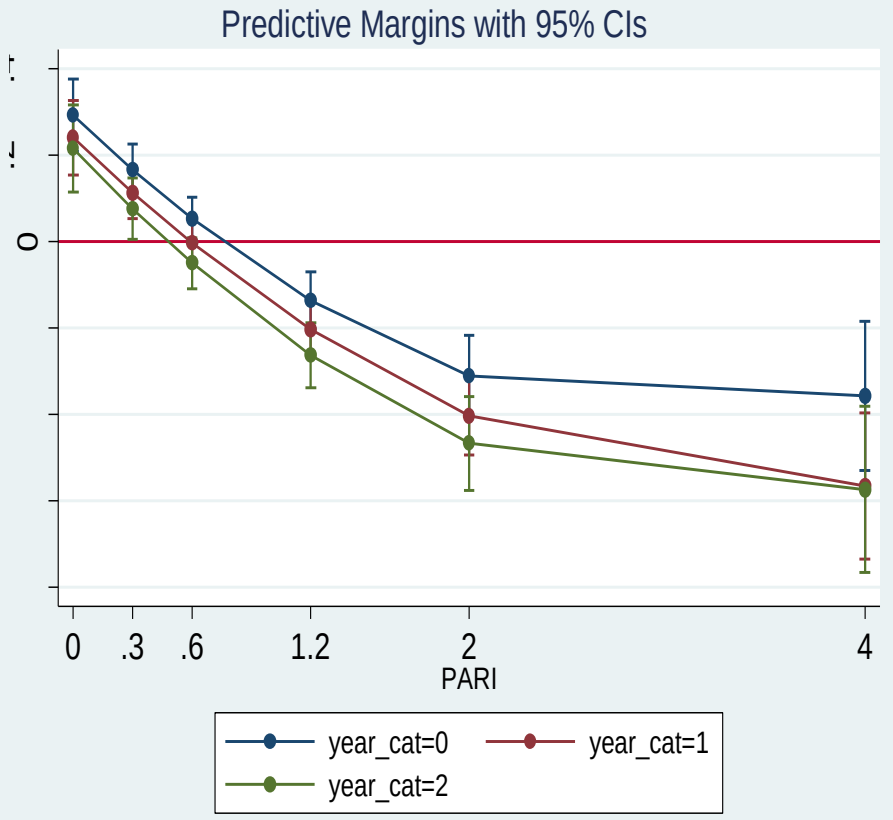
REMOTENESS

Remoteness: English

MALES

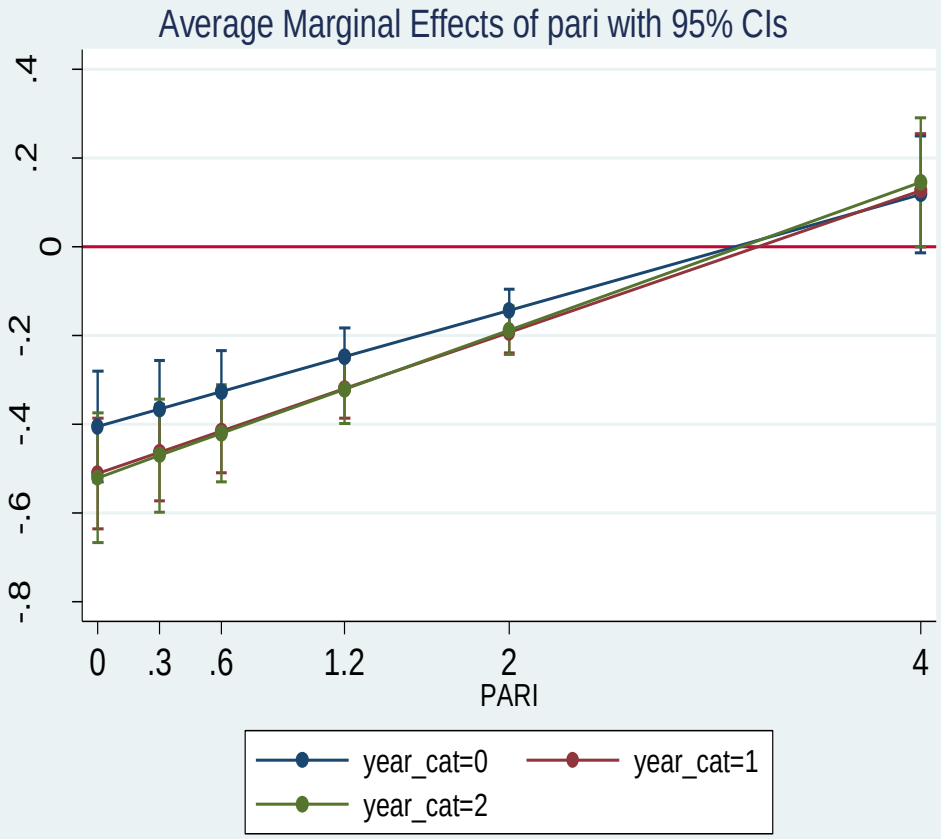


FEMALES

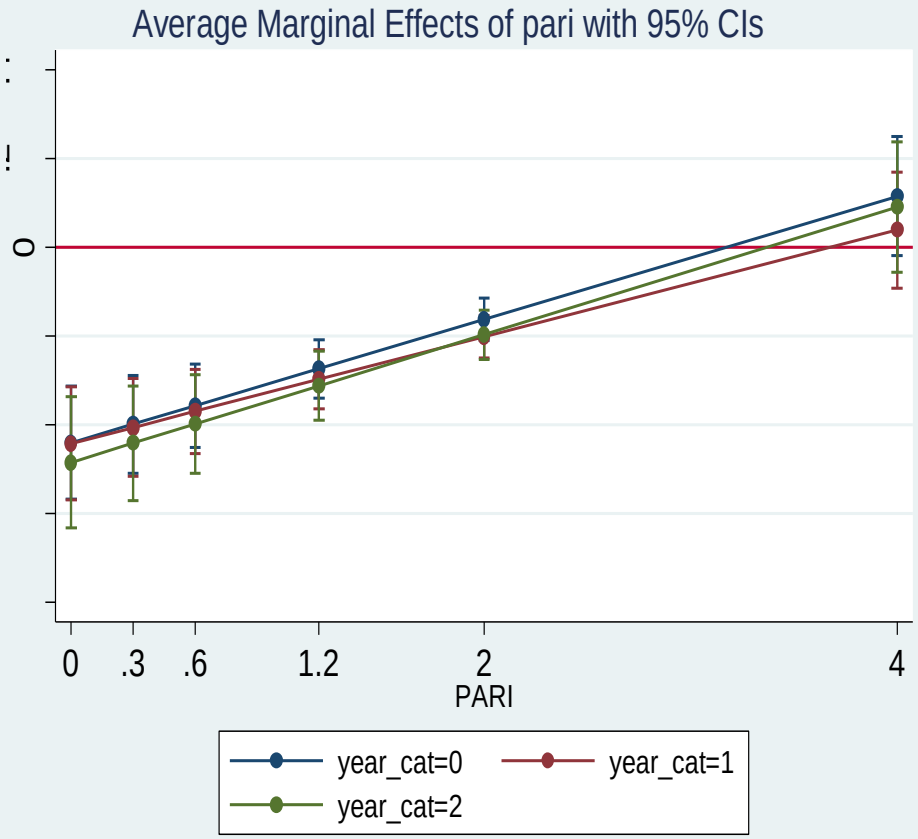


Remoteness: English

MALES

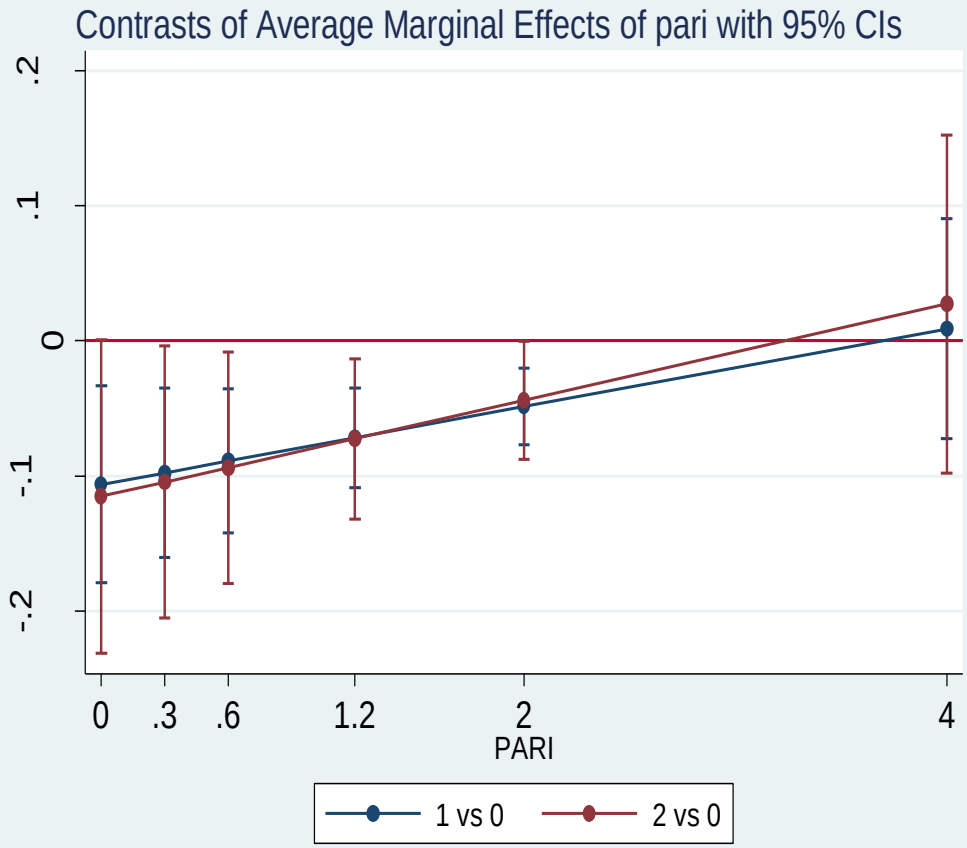


FEMALES

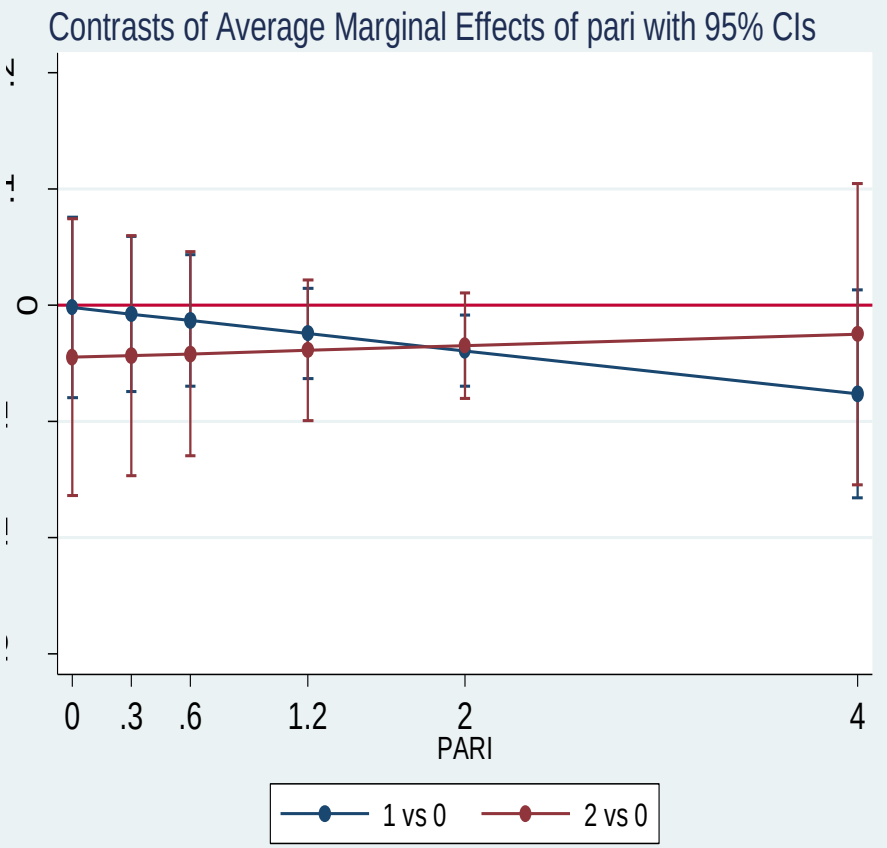


Remoteness: English

MALES

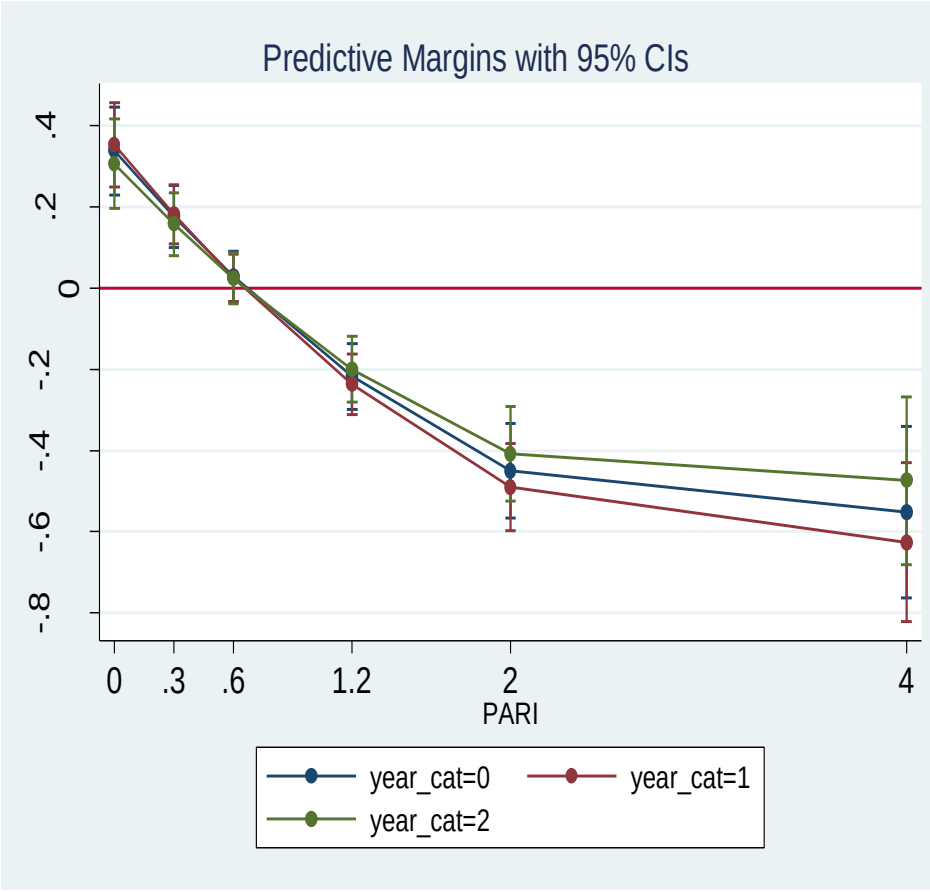


FEMALES

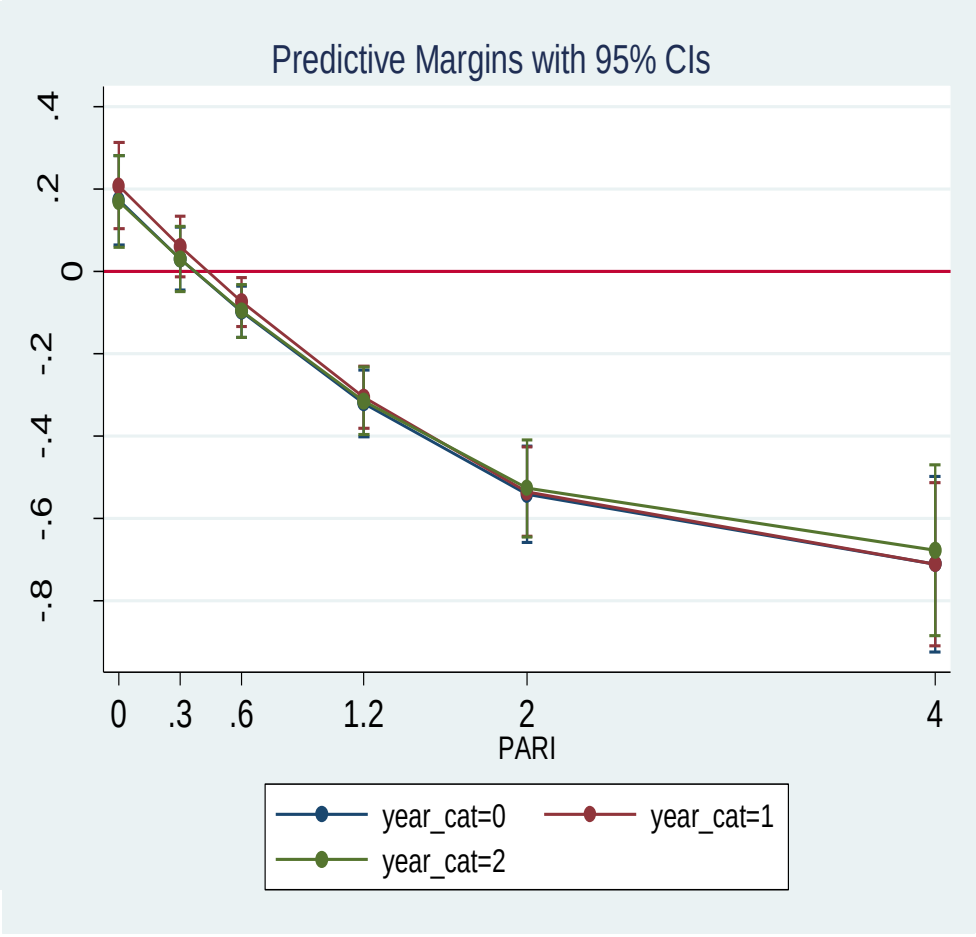


Remoteness: Maths

MALES

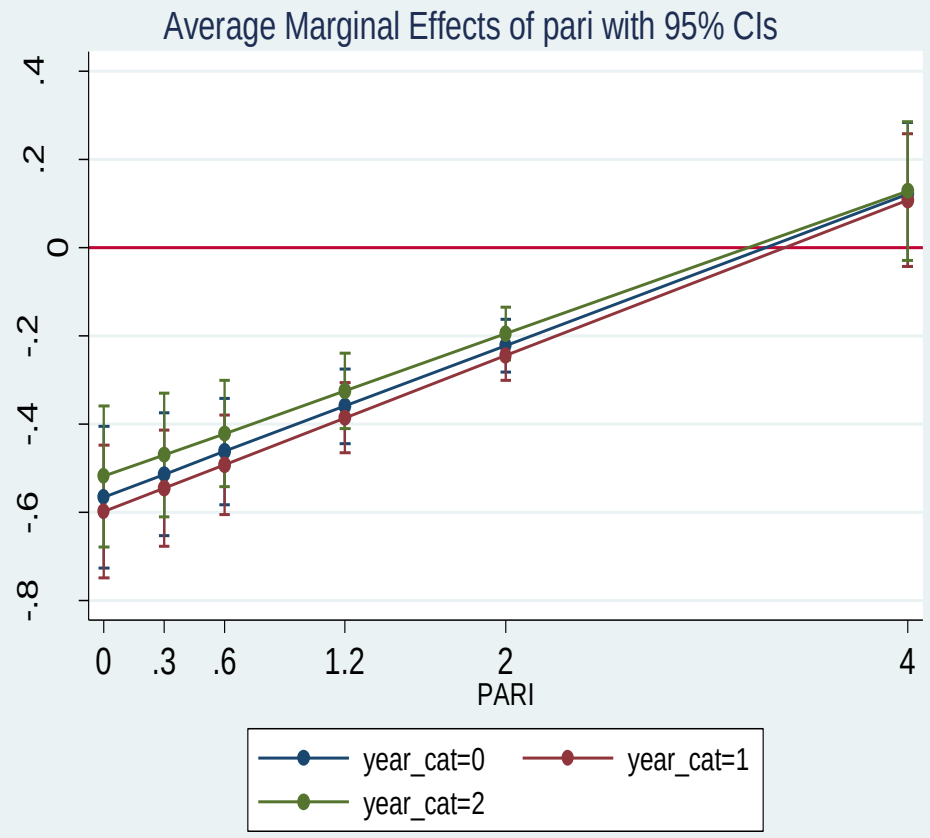


FEMALES

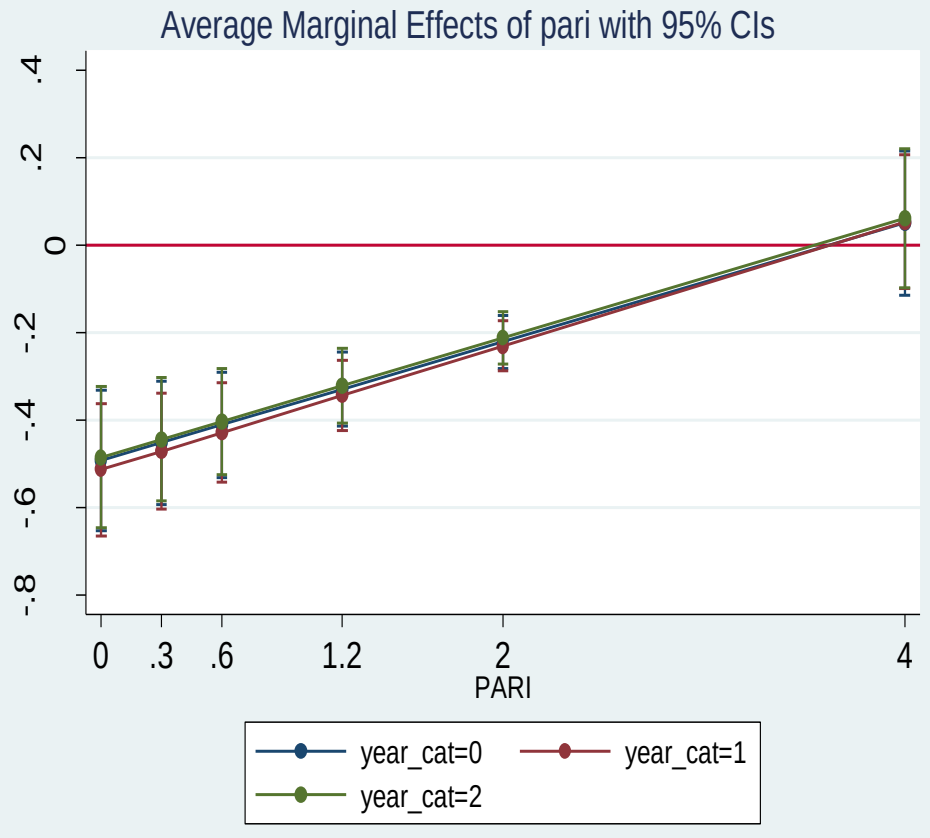


Remoteness: Maths

MALES

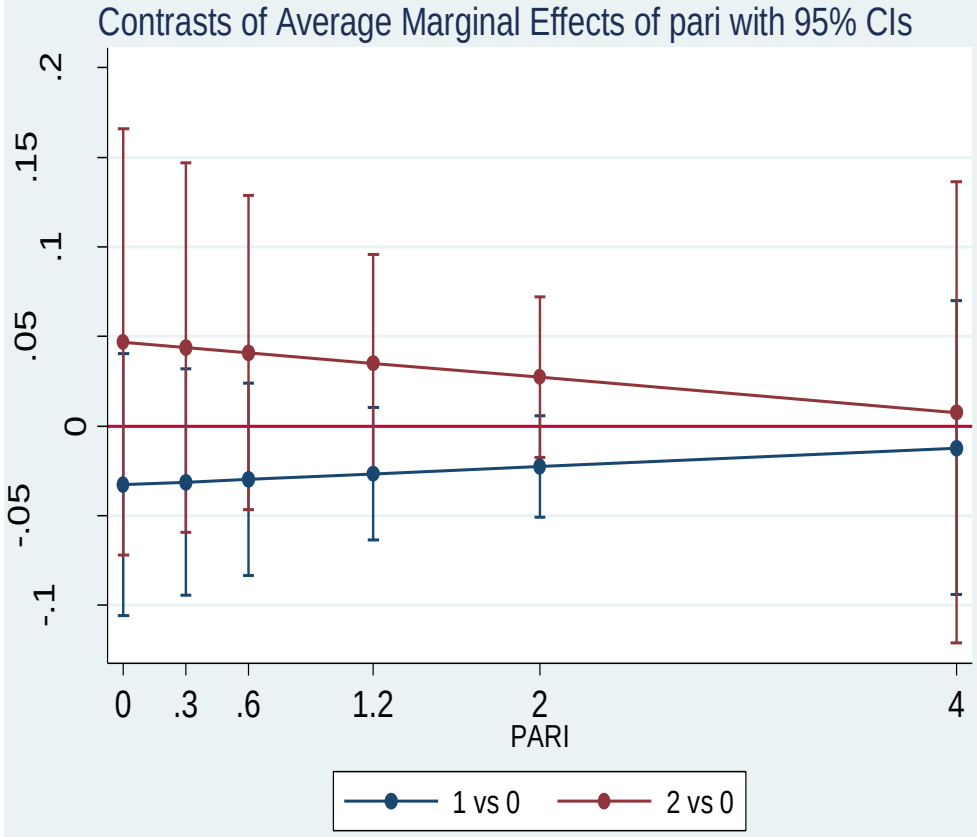


FEMALES

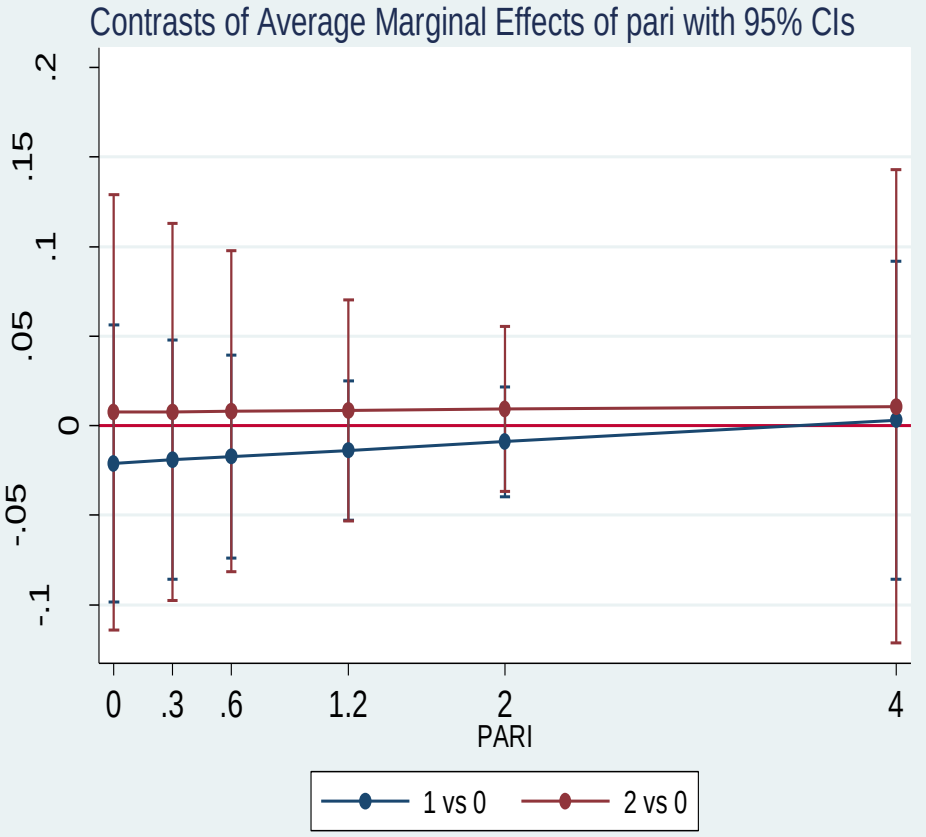


Remoteness: Maths

MALES



FEMALES



No significant changes in the average marginal effect over time

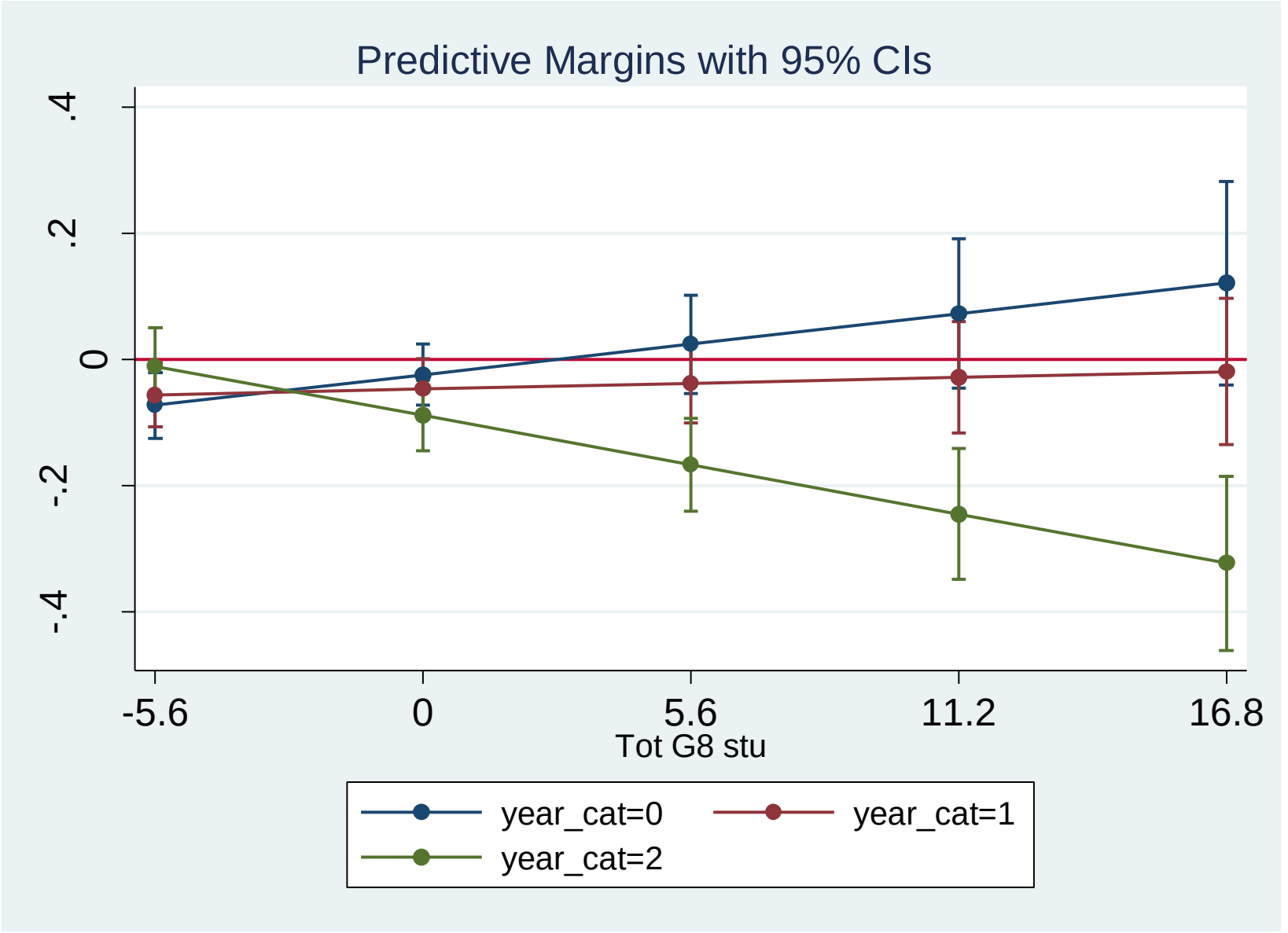
What explains remoteness effects over time?

- The negative effect of remoteness becomes worse with the introduction of the TFF for males in less remote areas (particularly English).
- Perhaps schools in urban areas are better able to use TFF funding from govt.
  - Remote schools need to access funds from banks which are difficult, expensive and time consuming to access for many schools.
  - Cost of transporting goods to rural areas.

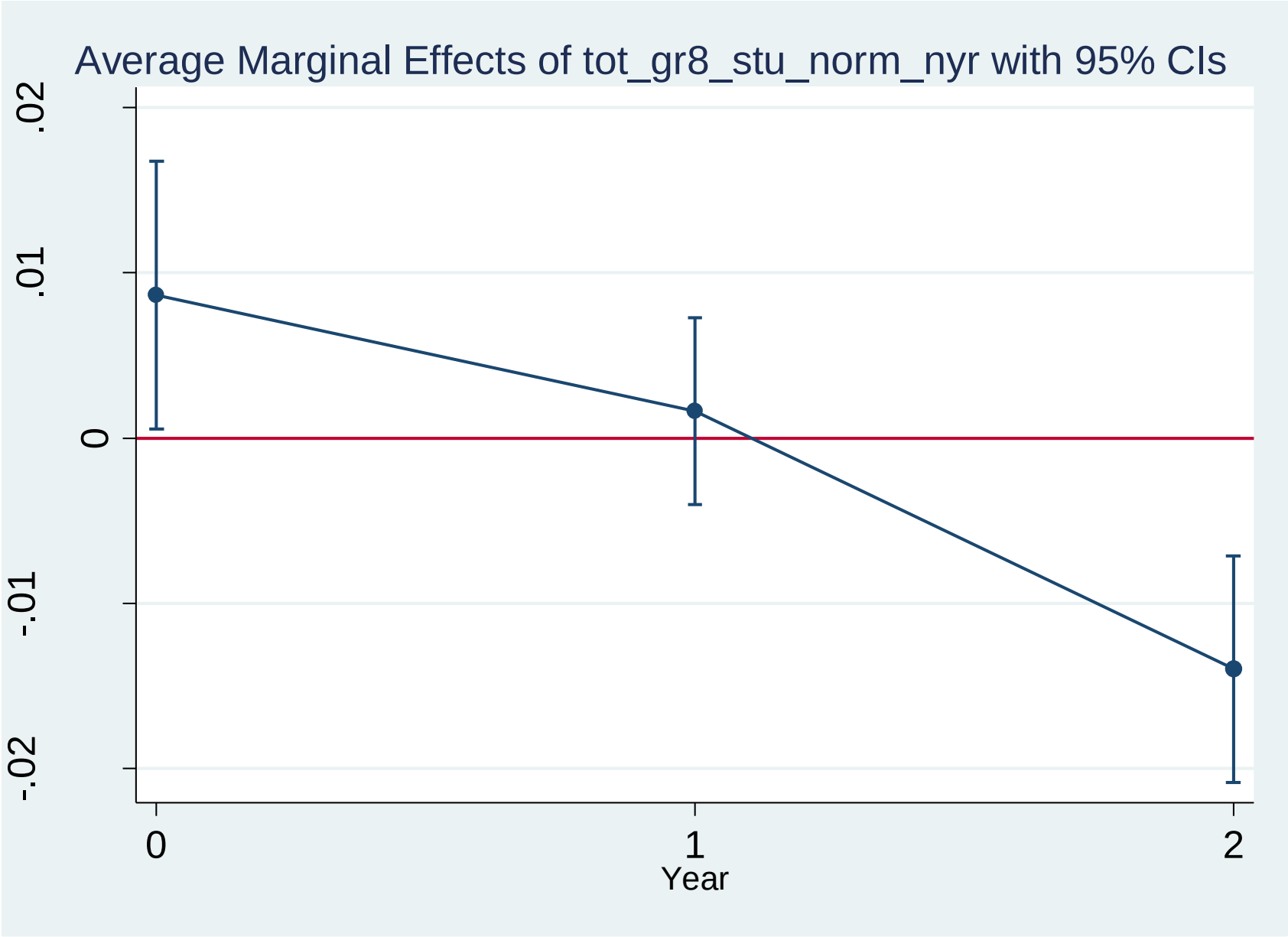
Results

SCHOOL SIZE (PER 10 GRADE 8 STUDENTS)

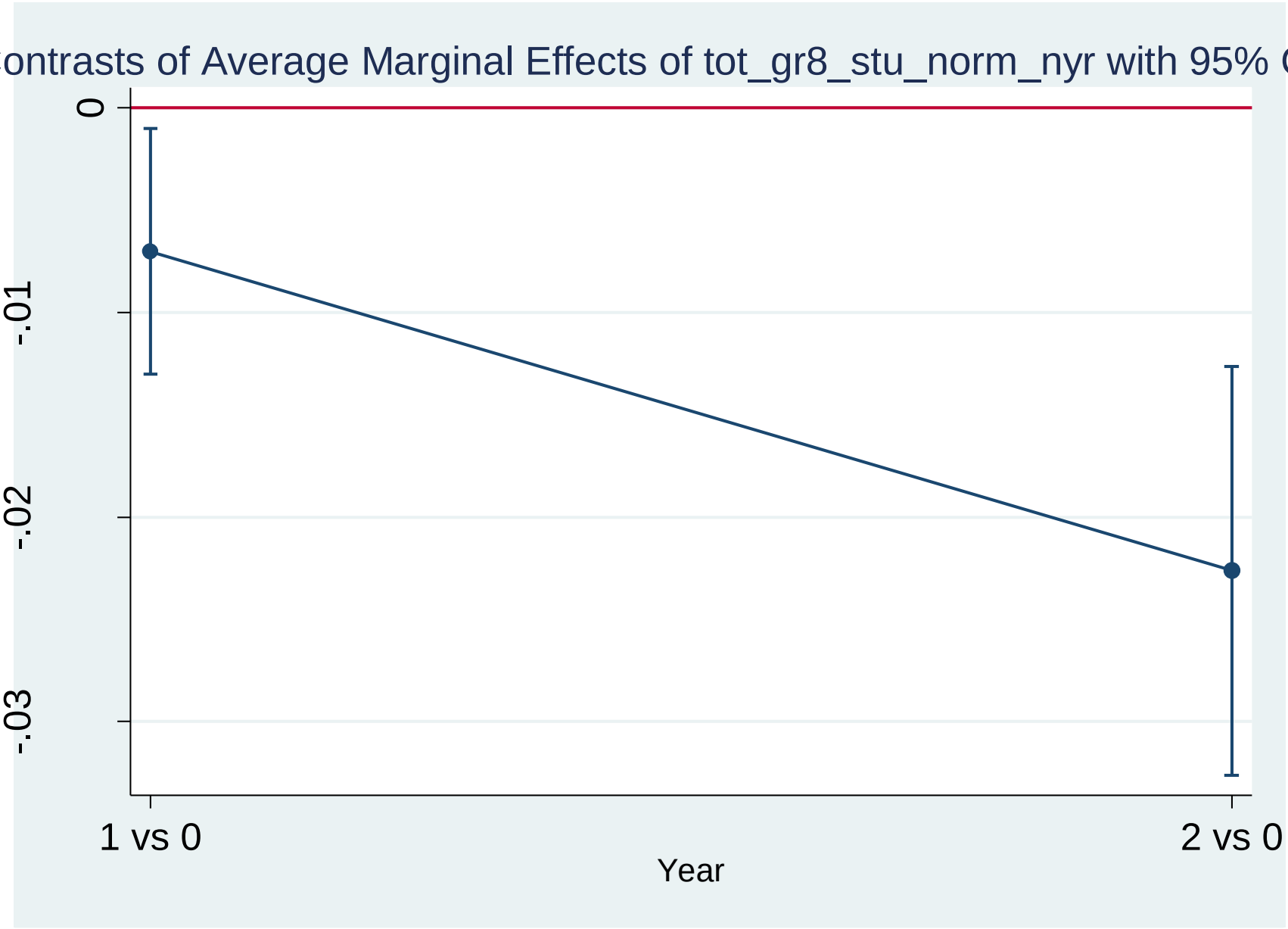
School size: English



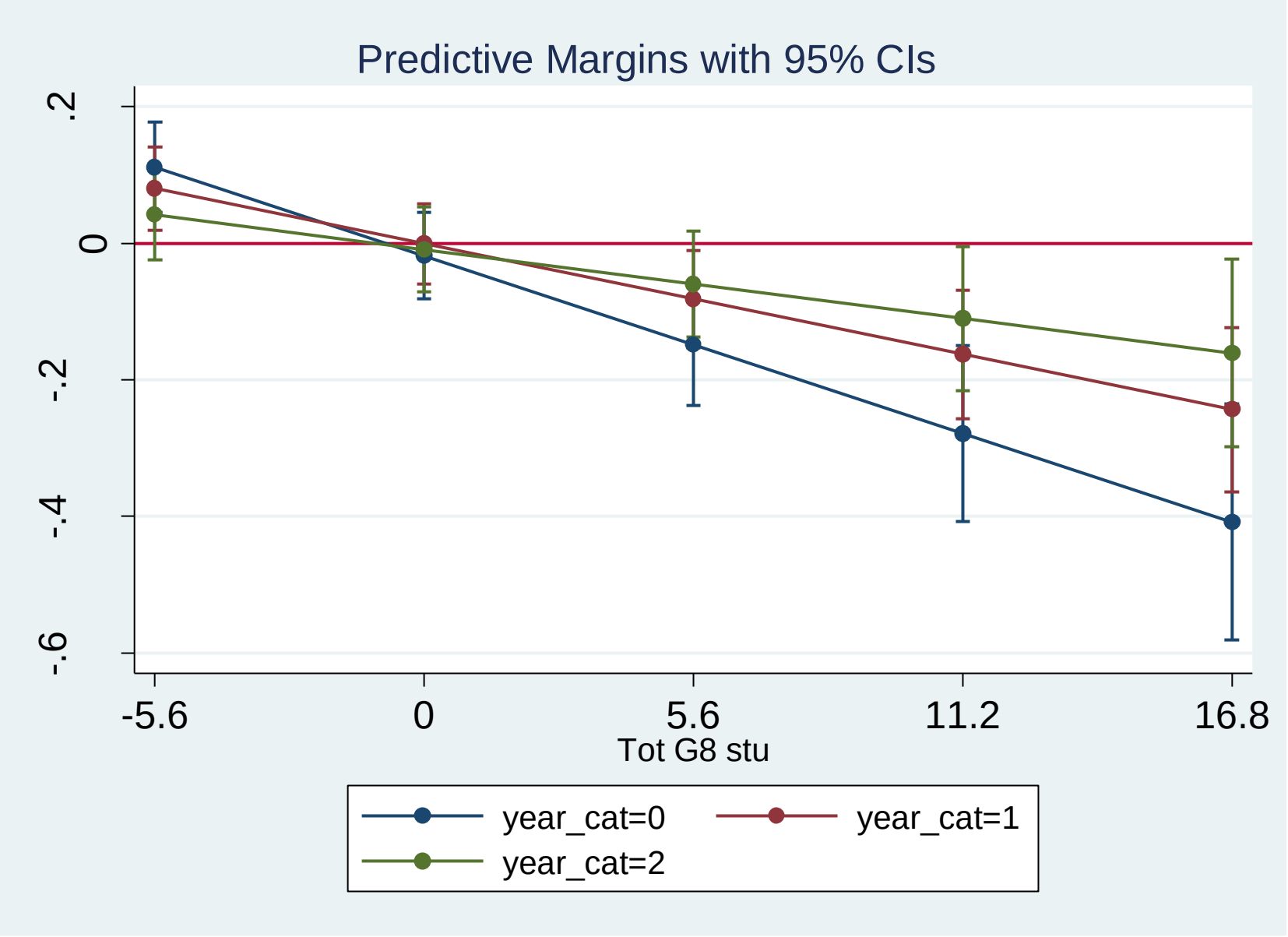
School size: English



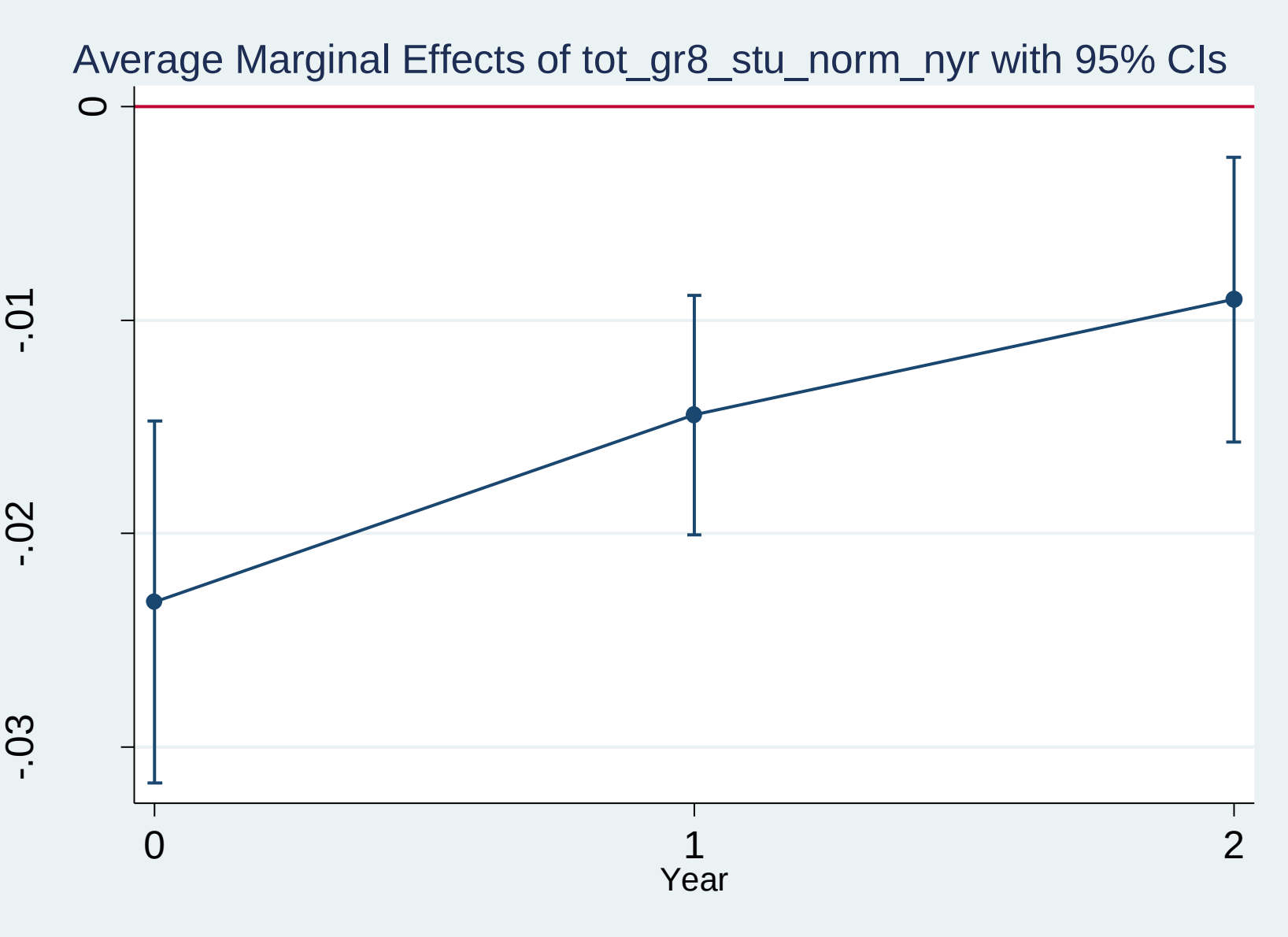
School size: English



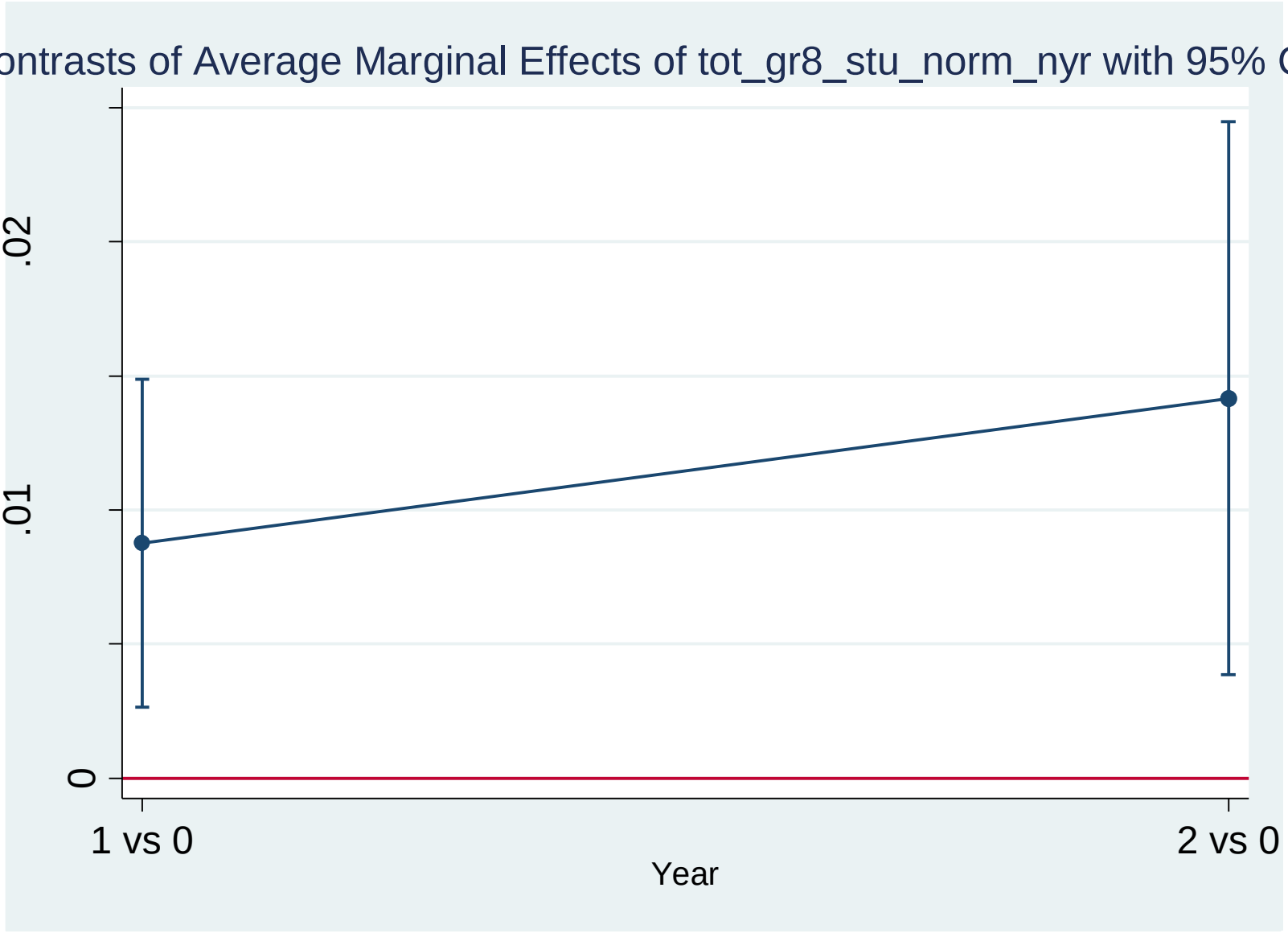
School size: Maths



School size: Maths



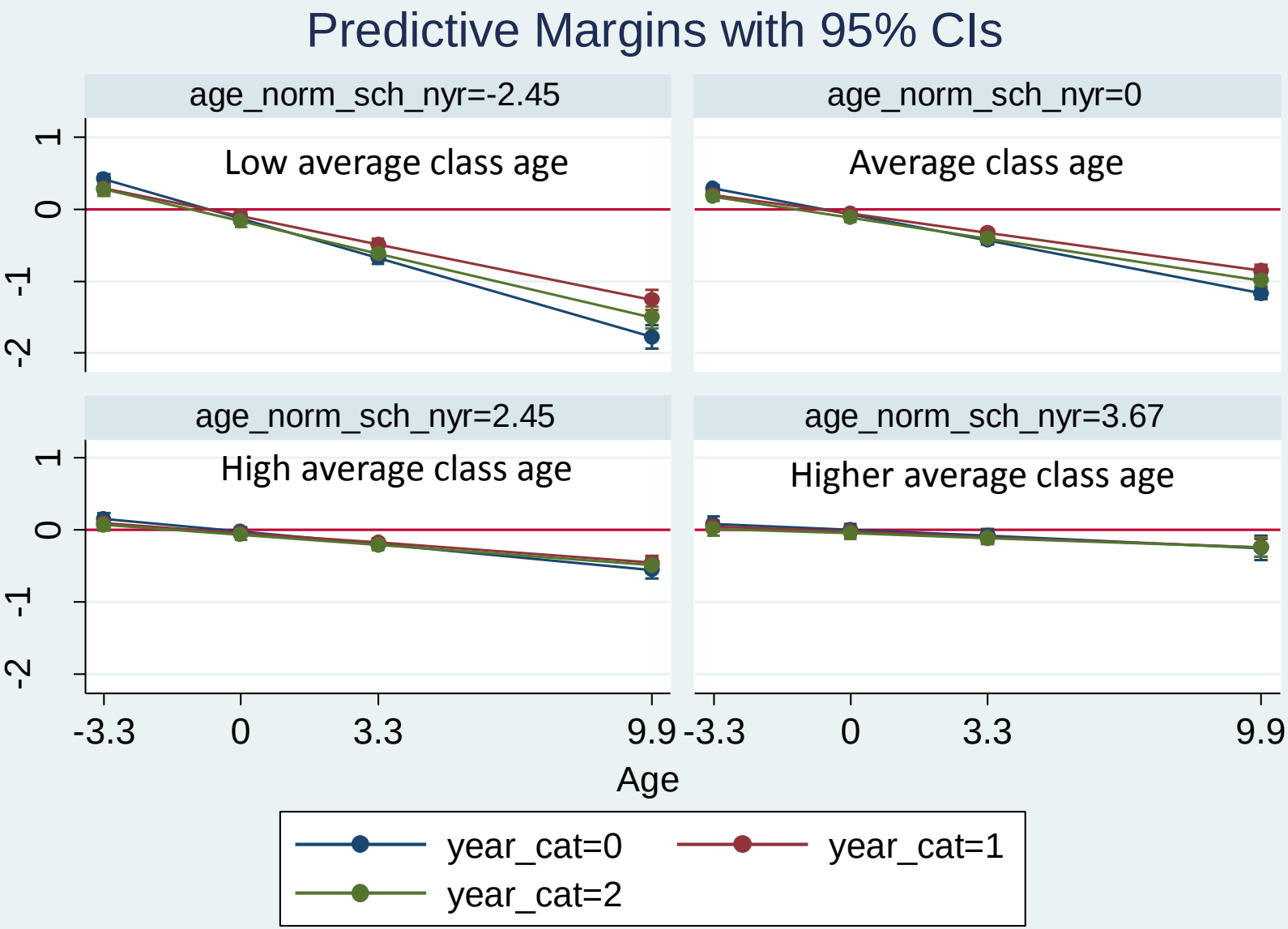
School size: Maths



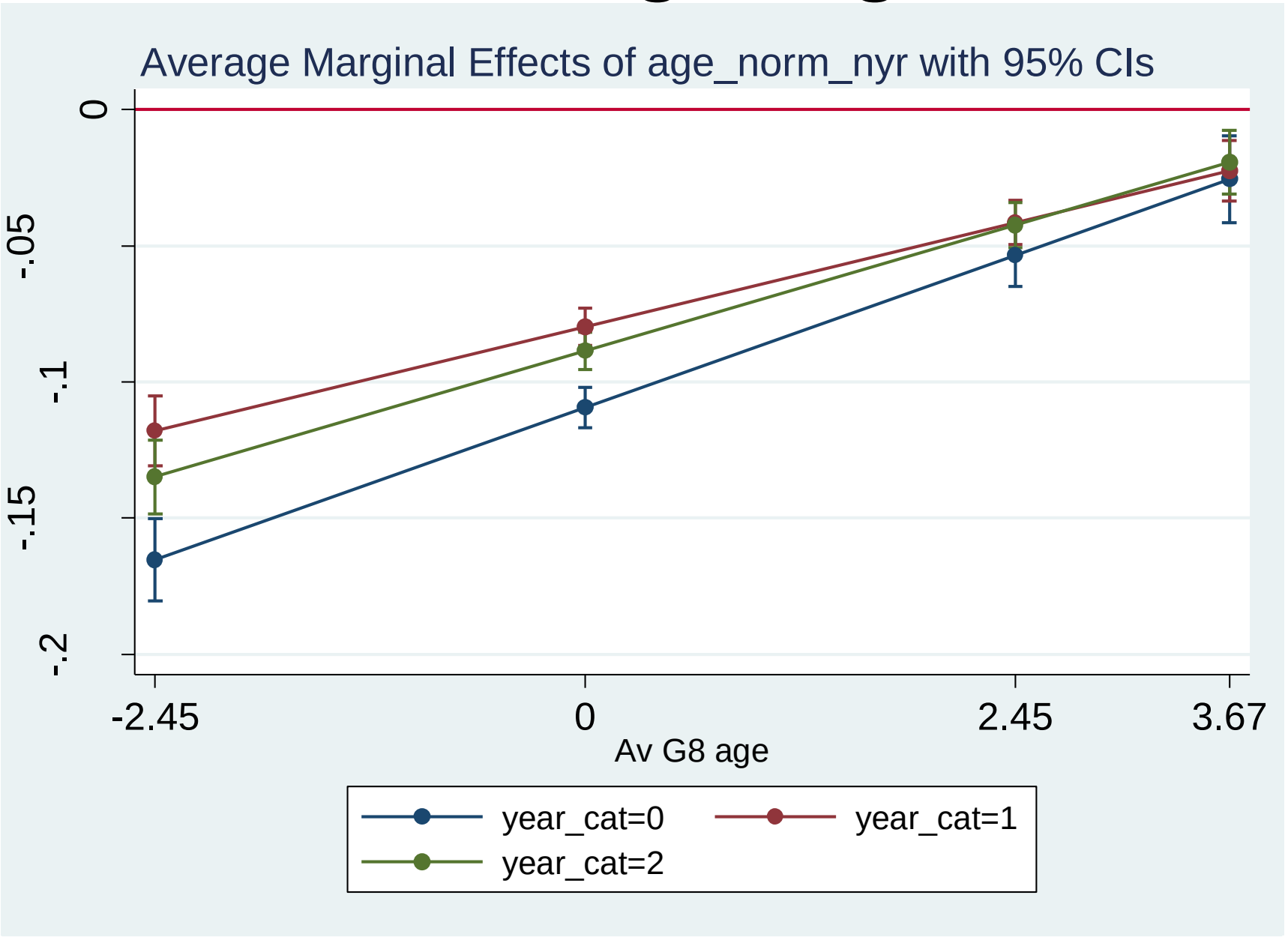
Results

STUDENT AGE

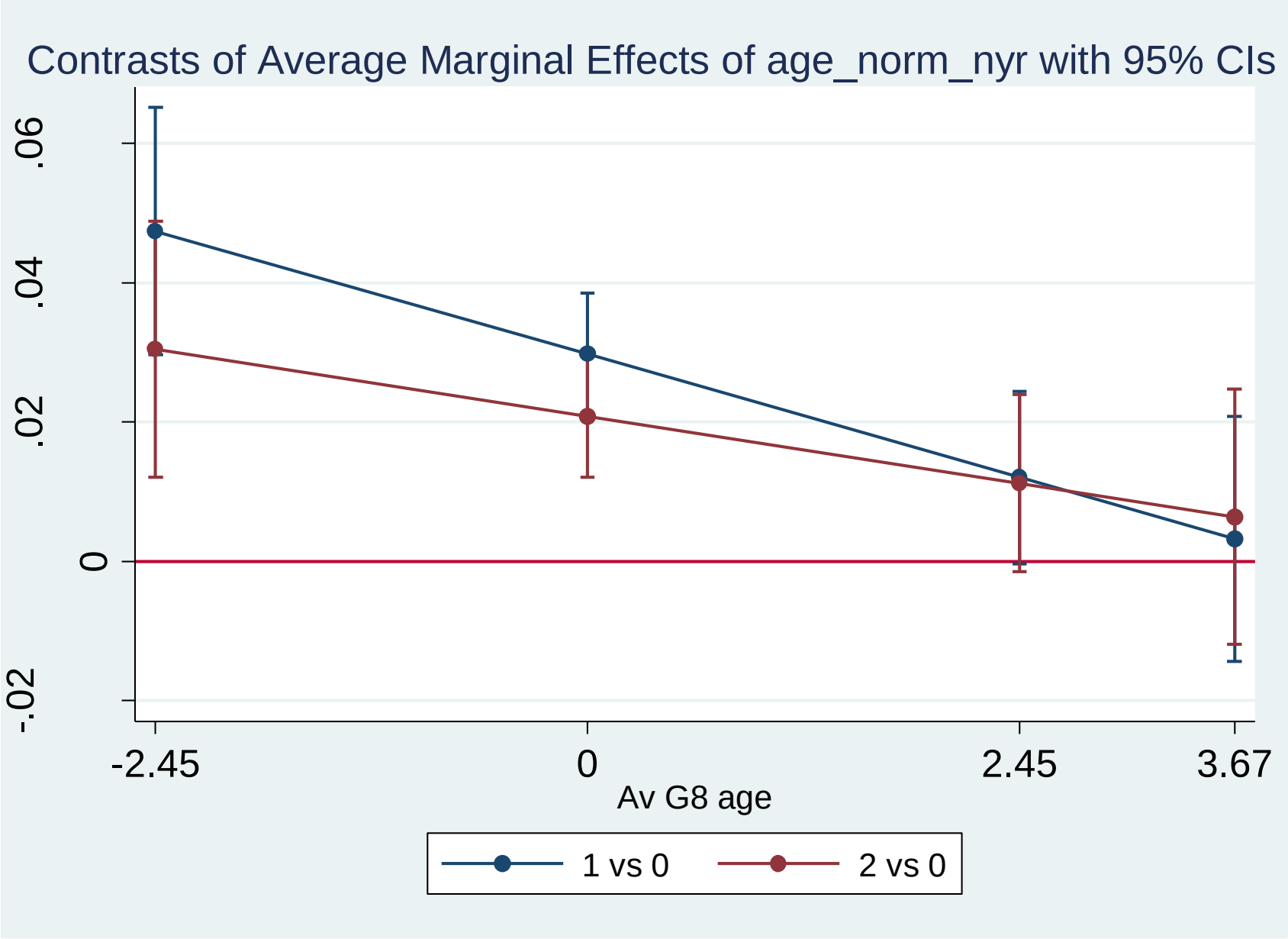
Student age: English



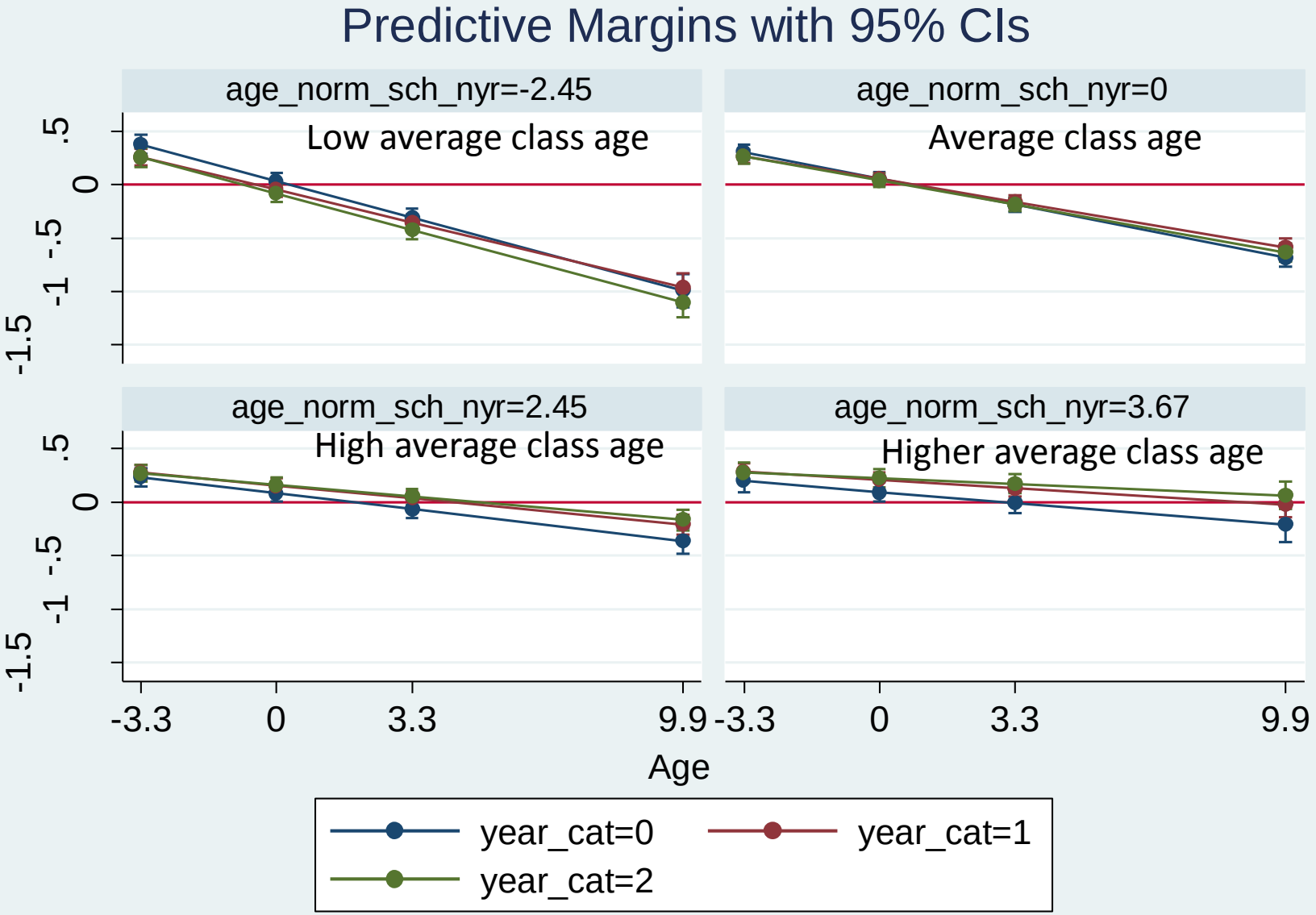
Student age: English



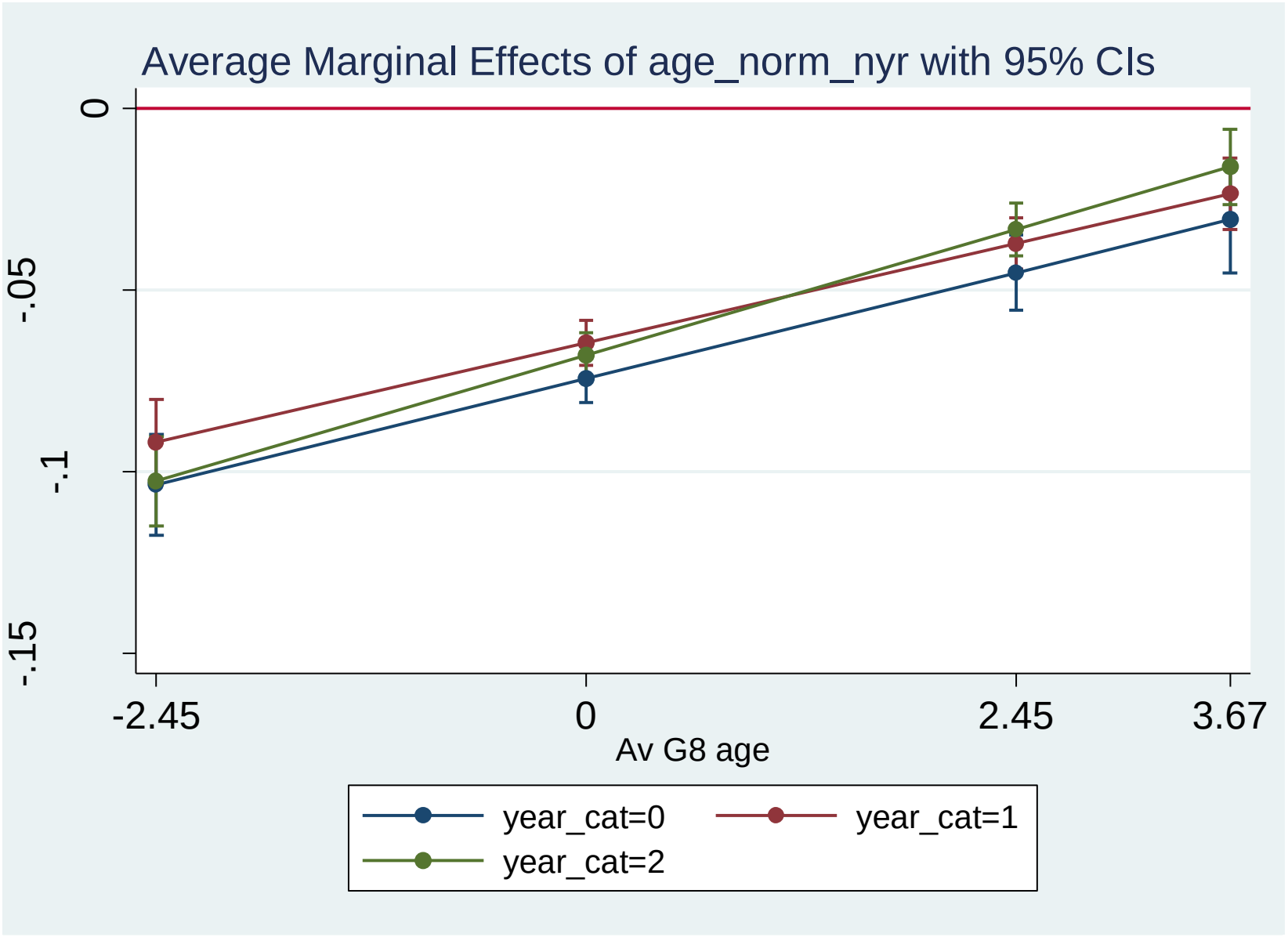
Student age: English



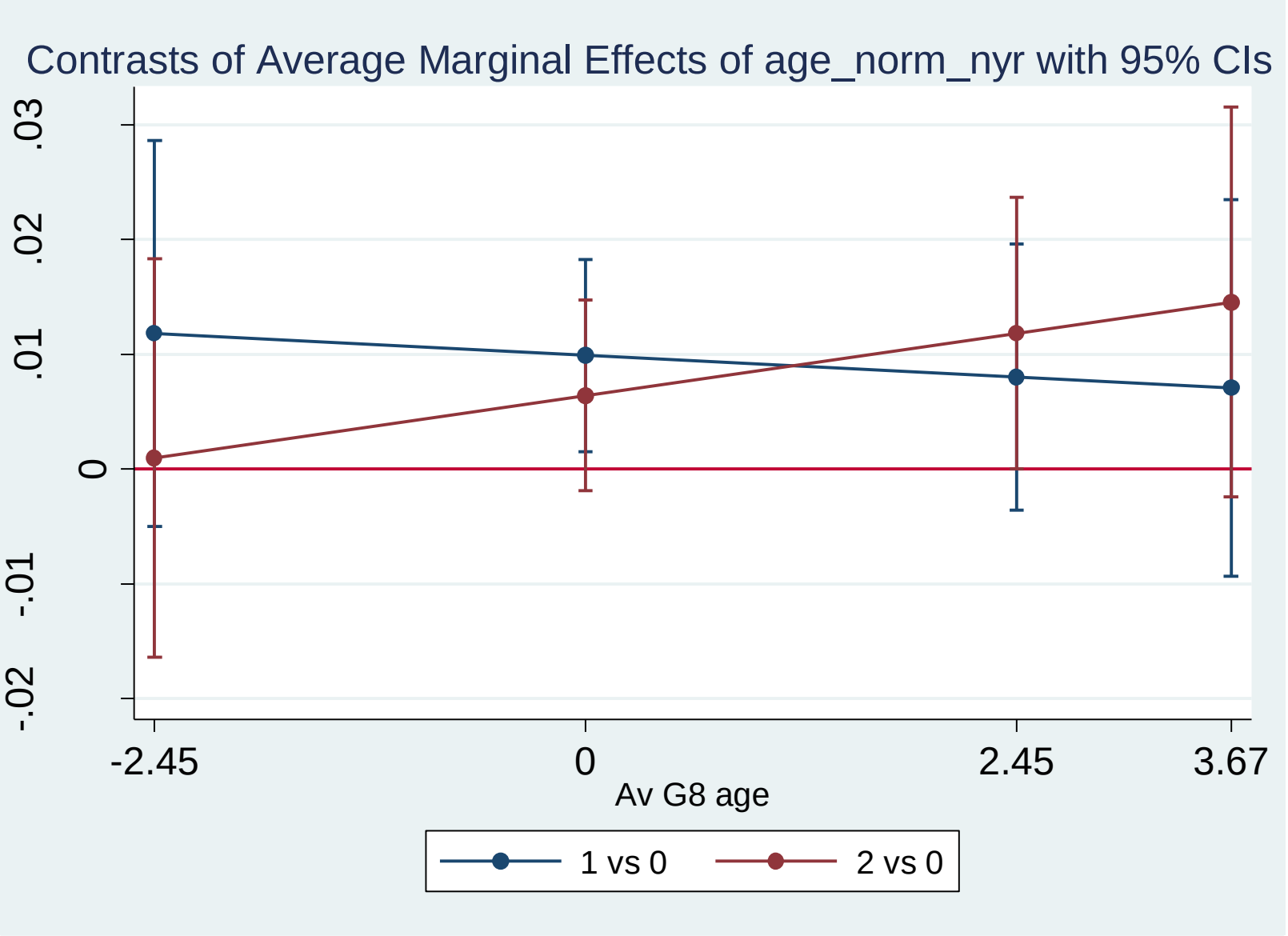
Student age: Maths



Student age: Maths



Student age: Maths



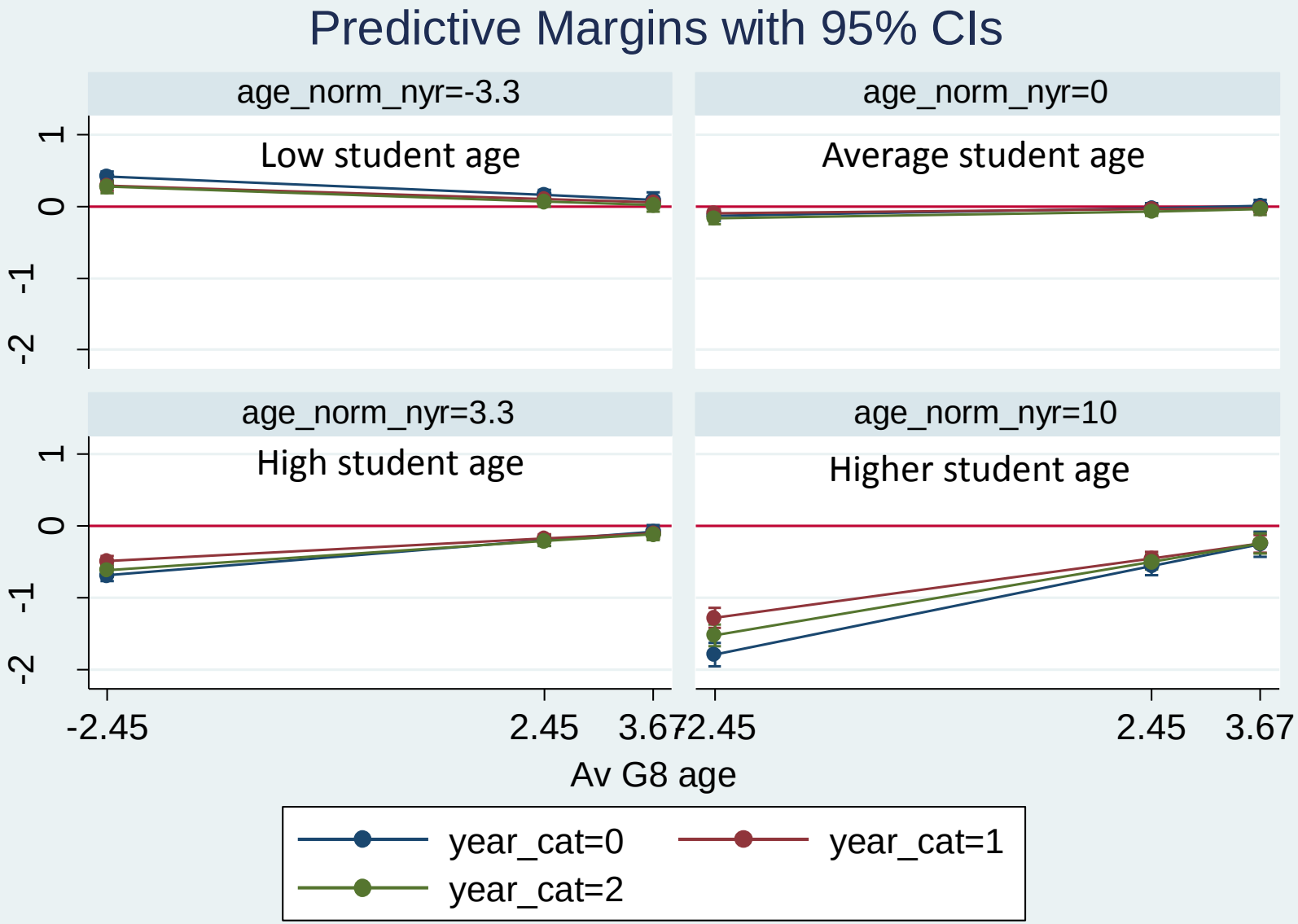
What might explain age effects?

- Age has a strong negative impact on test scores but the effect is moderated by the average age of the class.
  - Older students do better in classes with a higher average age.
  - Age is also a proxy for SES and IQ.
- The introduction of the TFF is associated with a reduction in the negative effect of being older in a younger class.
- Possible adaptation to changing class demographics as older students return to school.

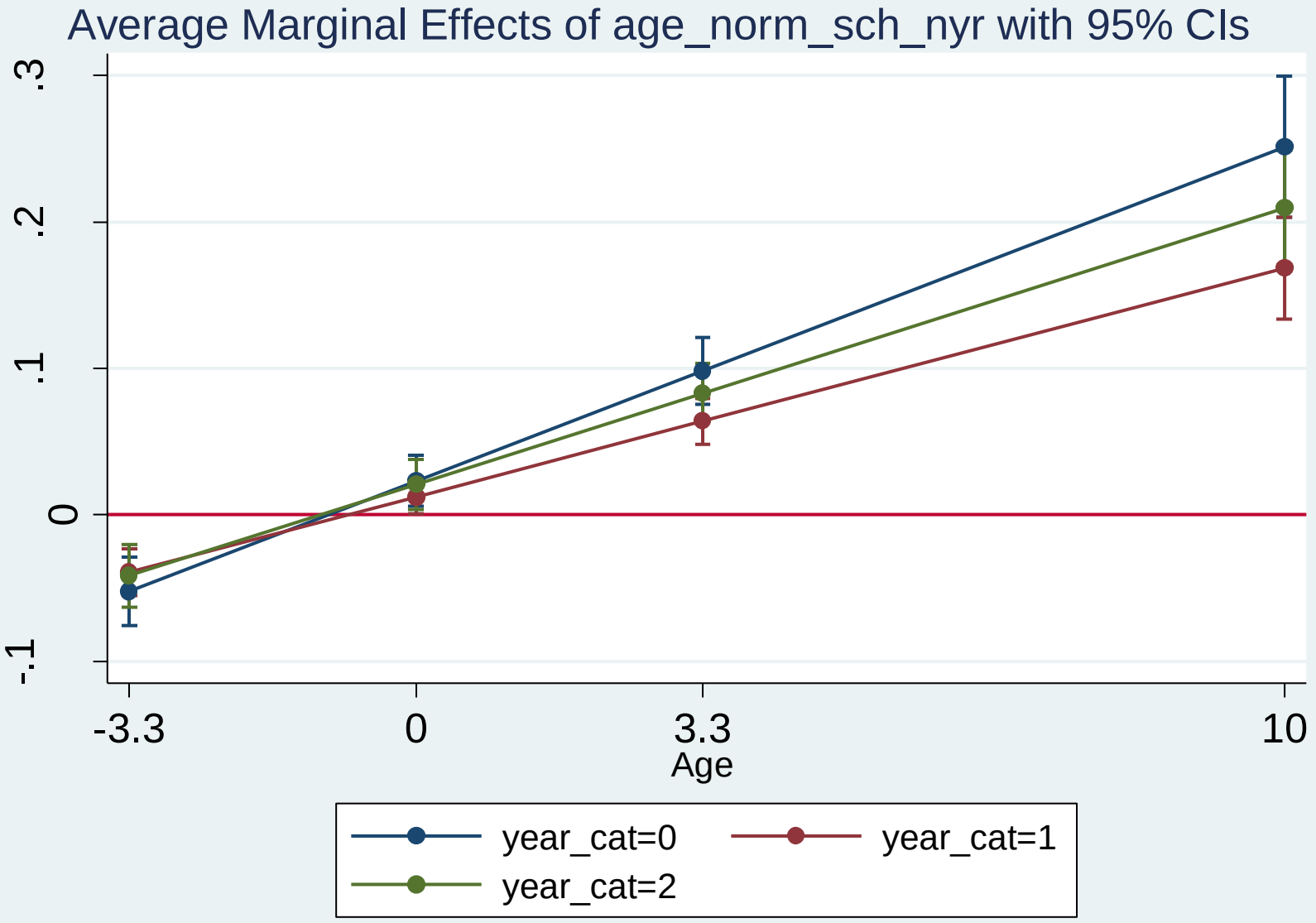
Results

AVERAGE CLASS AGE

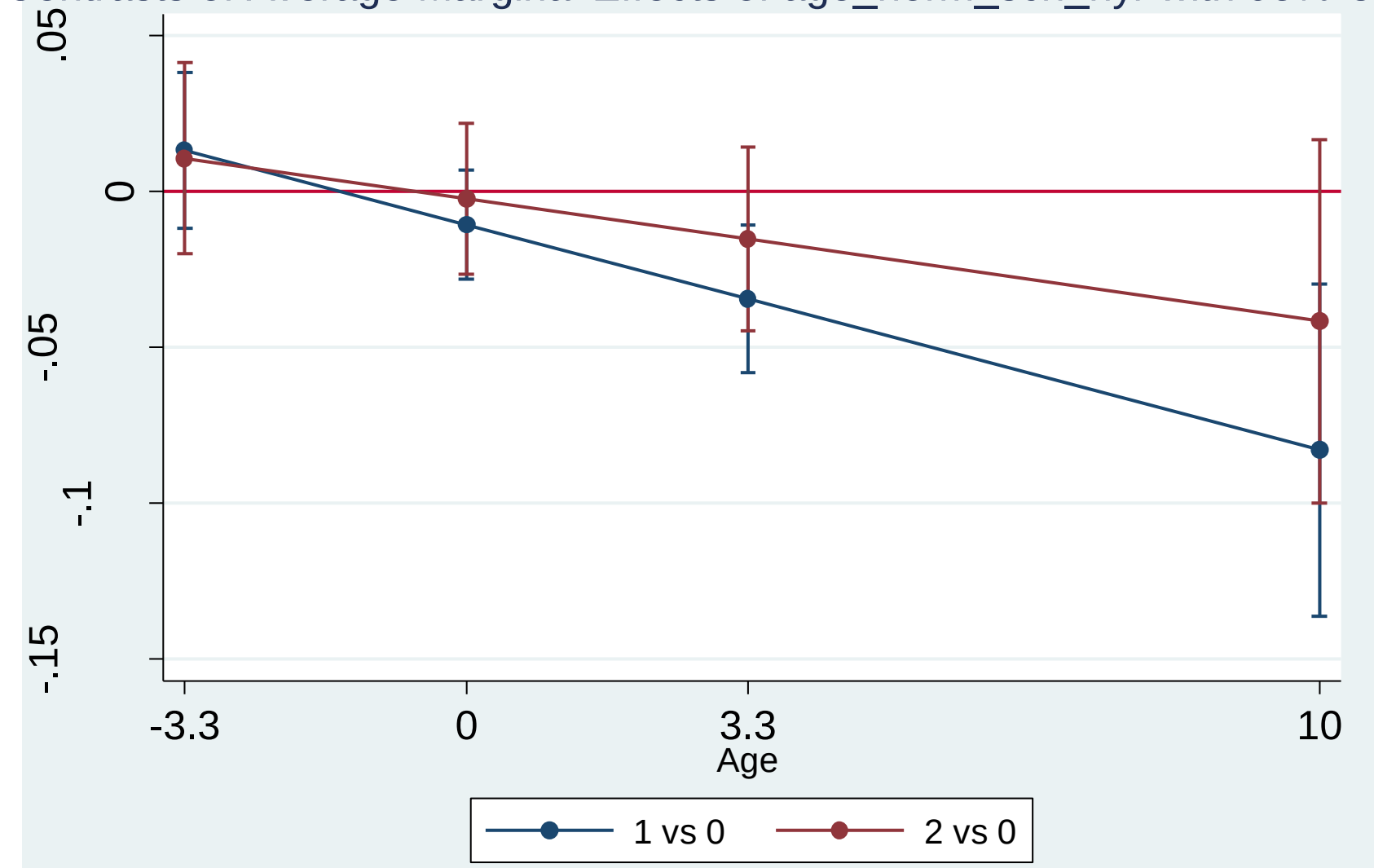
Class age: English



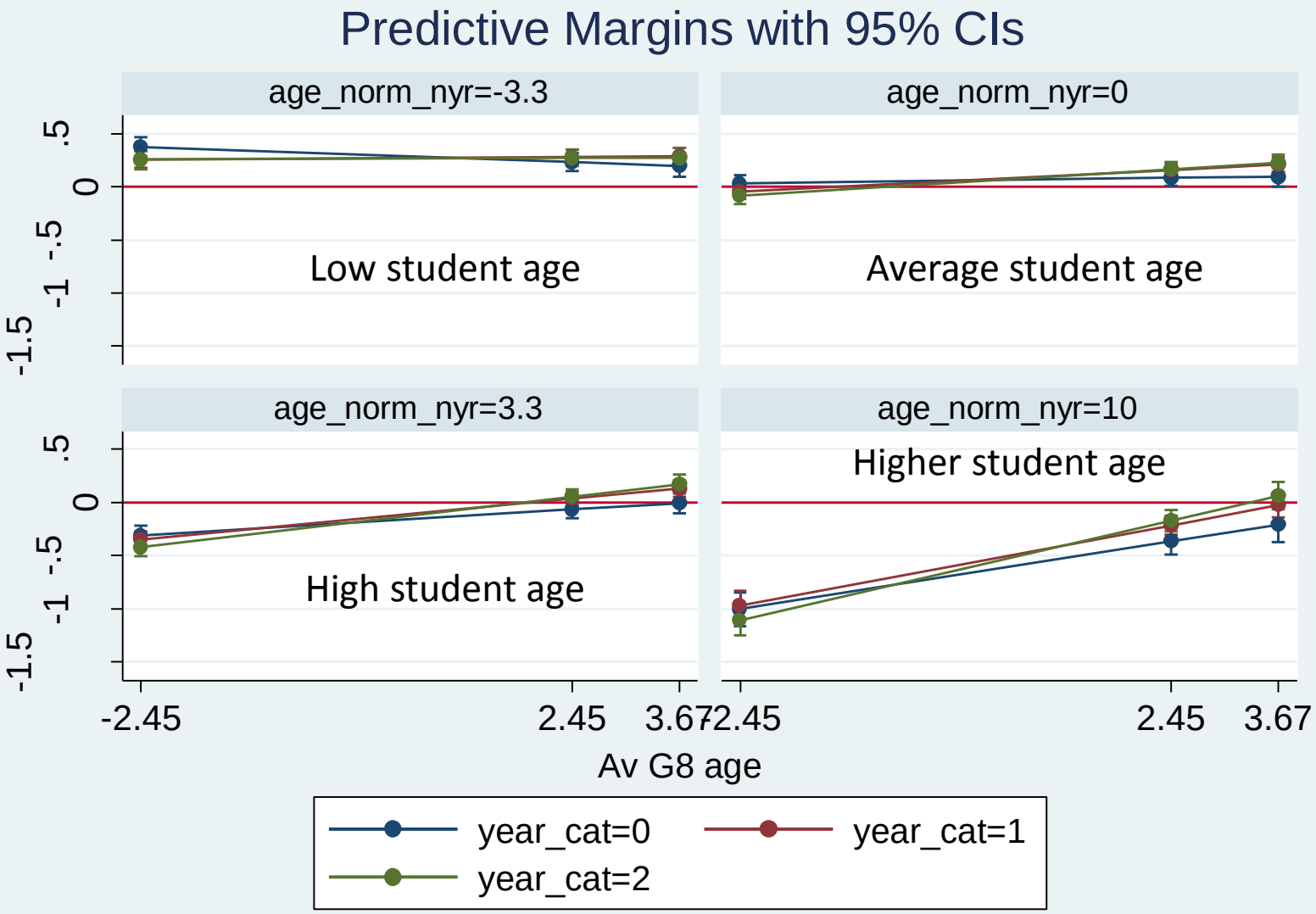
Class age: English



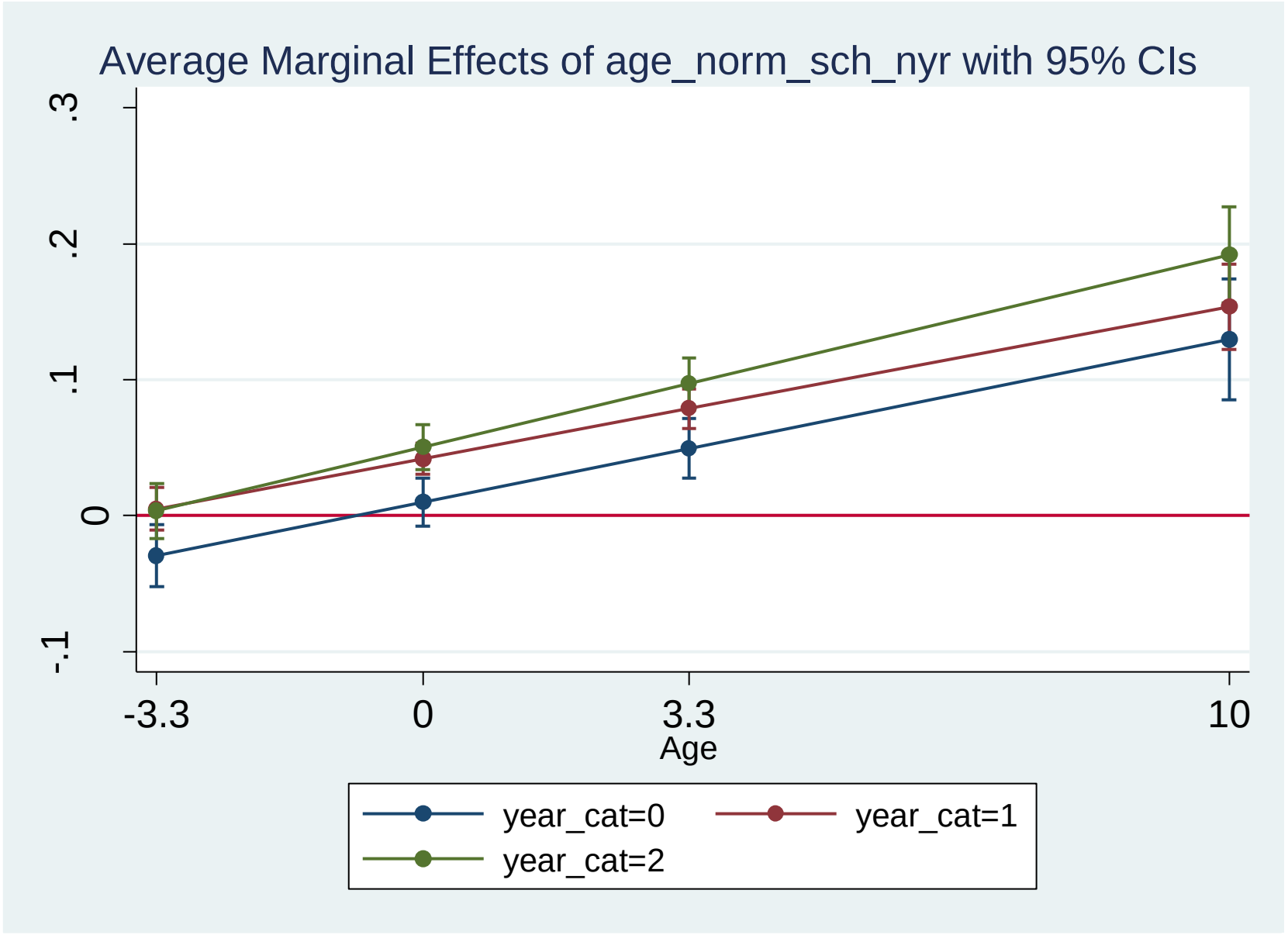
Class age: English



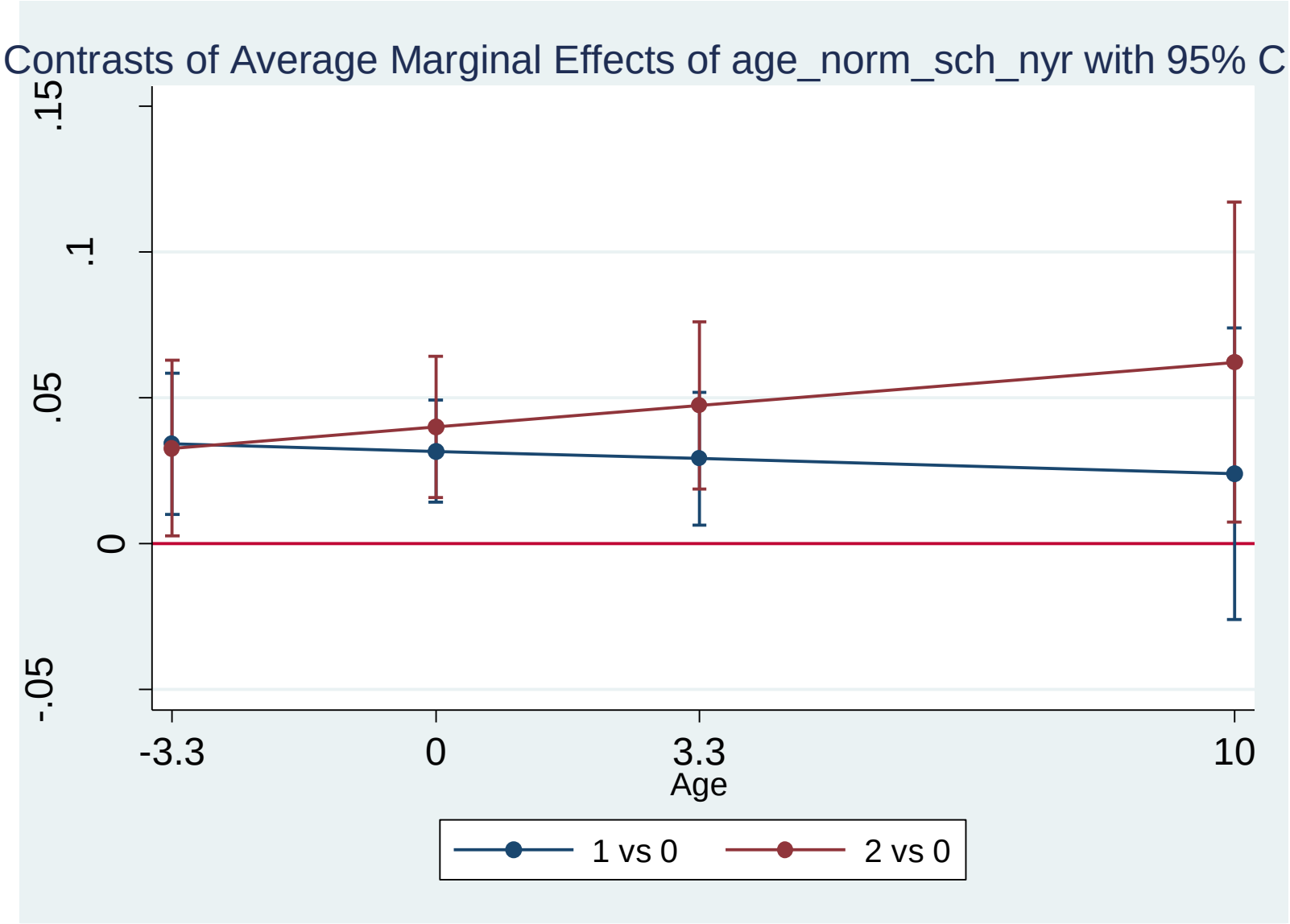
Class age: Maths



Class age: Maths



Class age: Maths



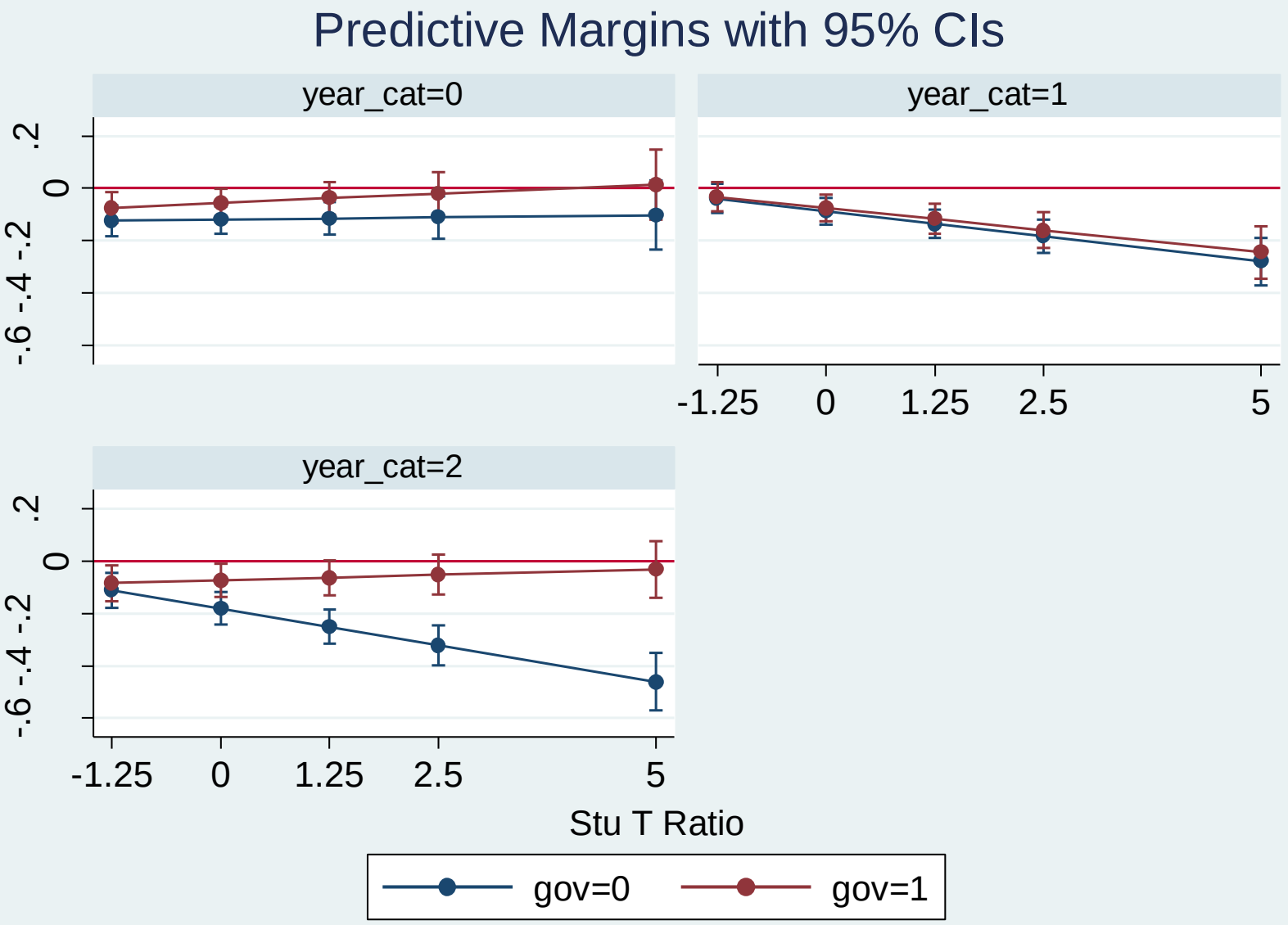
What might explain class average age effects?

- There are strong average age effects which work to benefit older students but negatively impact younger students.
- The removal of school fees will tend to increase the number students of relatively low ability / or low socioeconomic level. These new students will tend to be older. So an increase in the average age in itself may not have a direct effect on individual learning outcomes but instead through its association with other characteristics.
- However, in Melanesian culture, respect is related to age and so the learning outcomes of young students may be disrupted in the presence of older students who capture more attention of teachers.

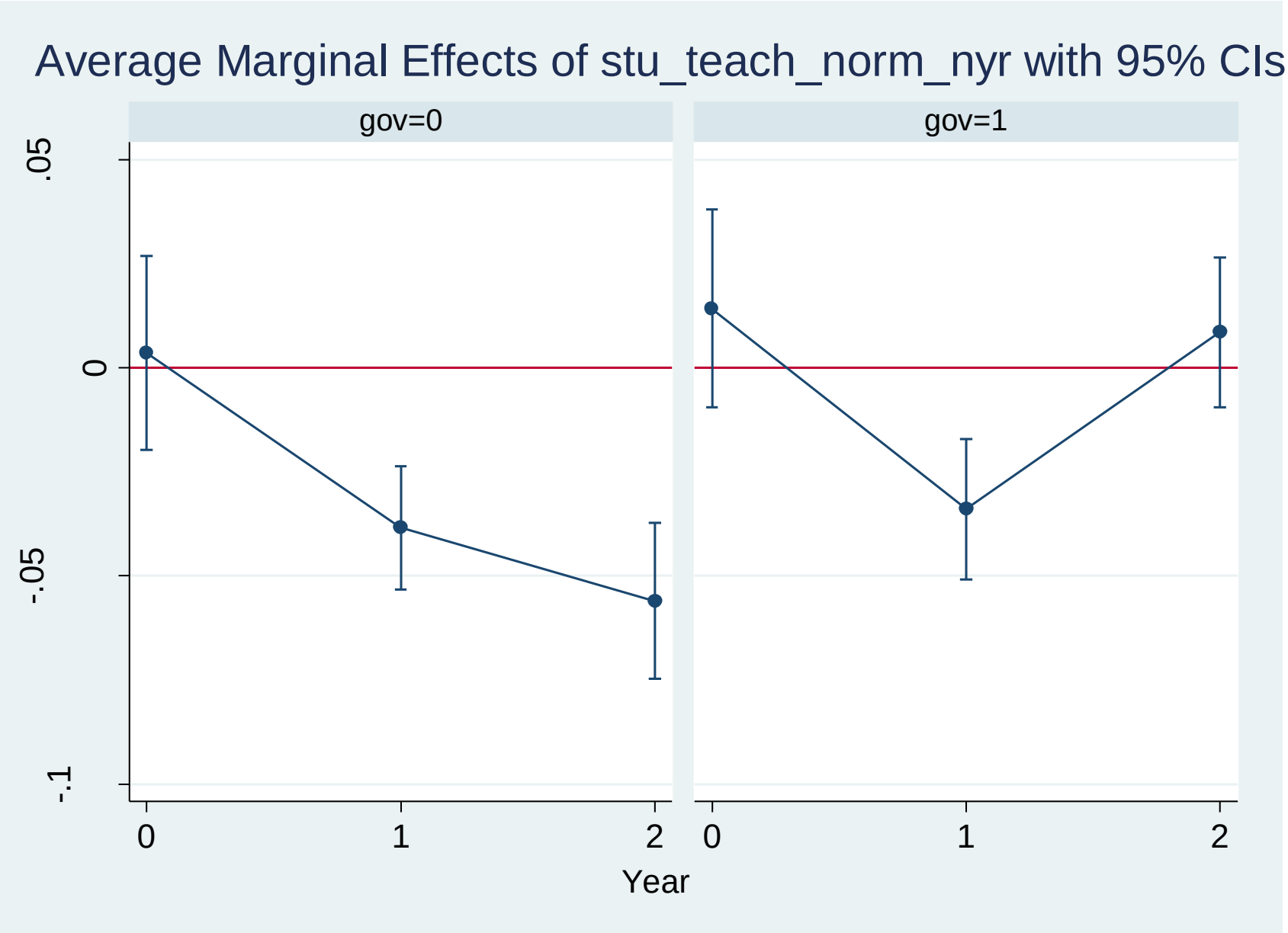
Results

CLASS SIZE (10 STUDENTS PER TEACHER)

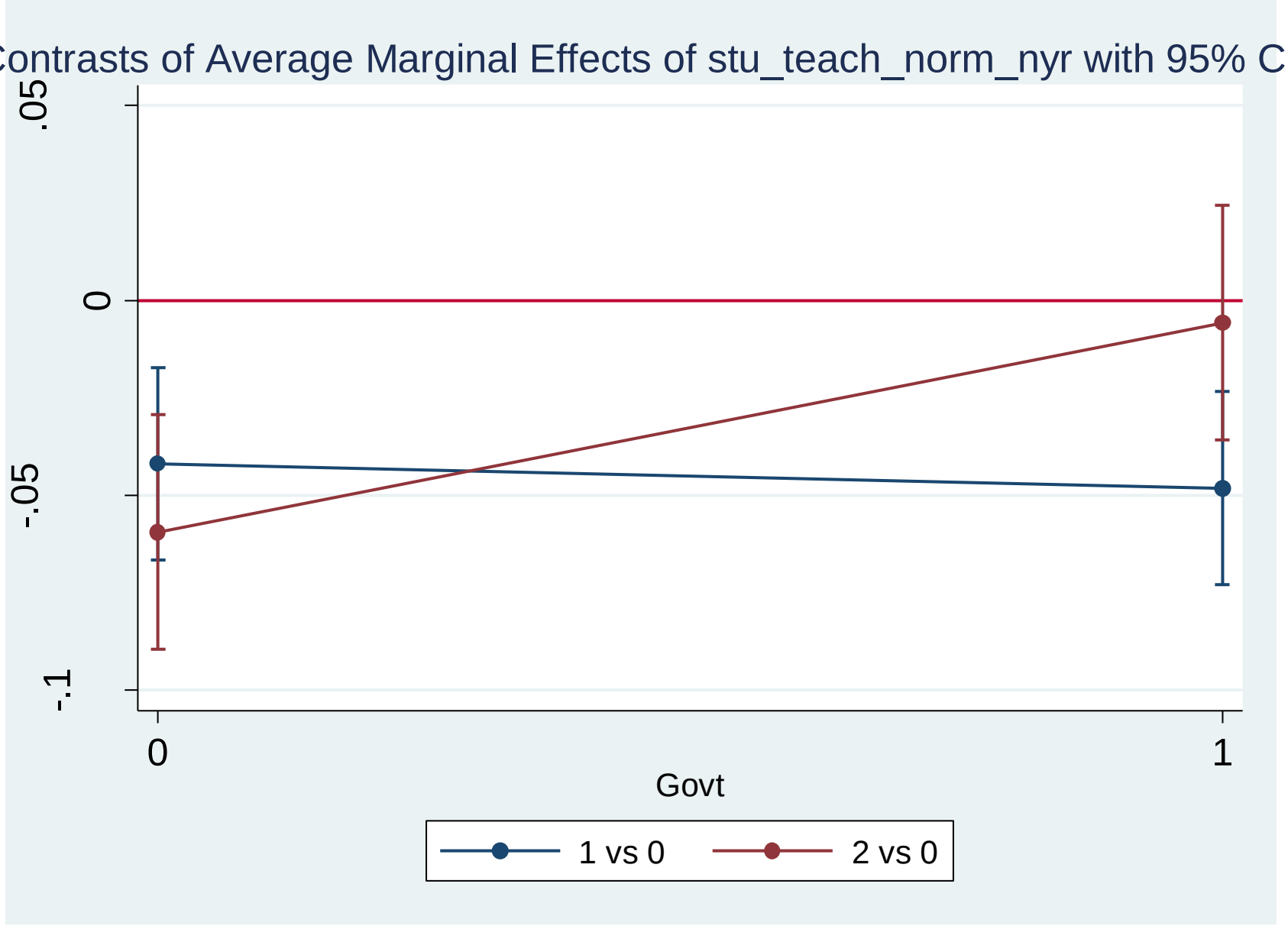
Class size: English



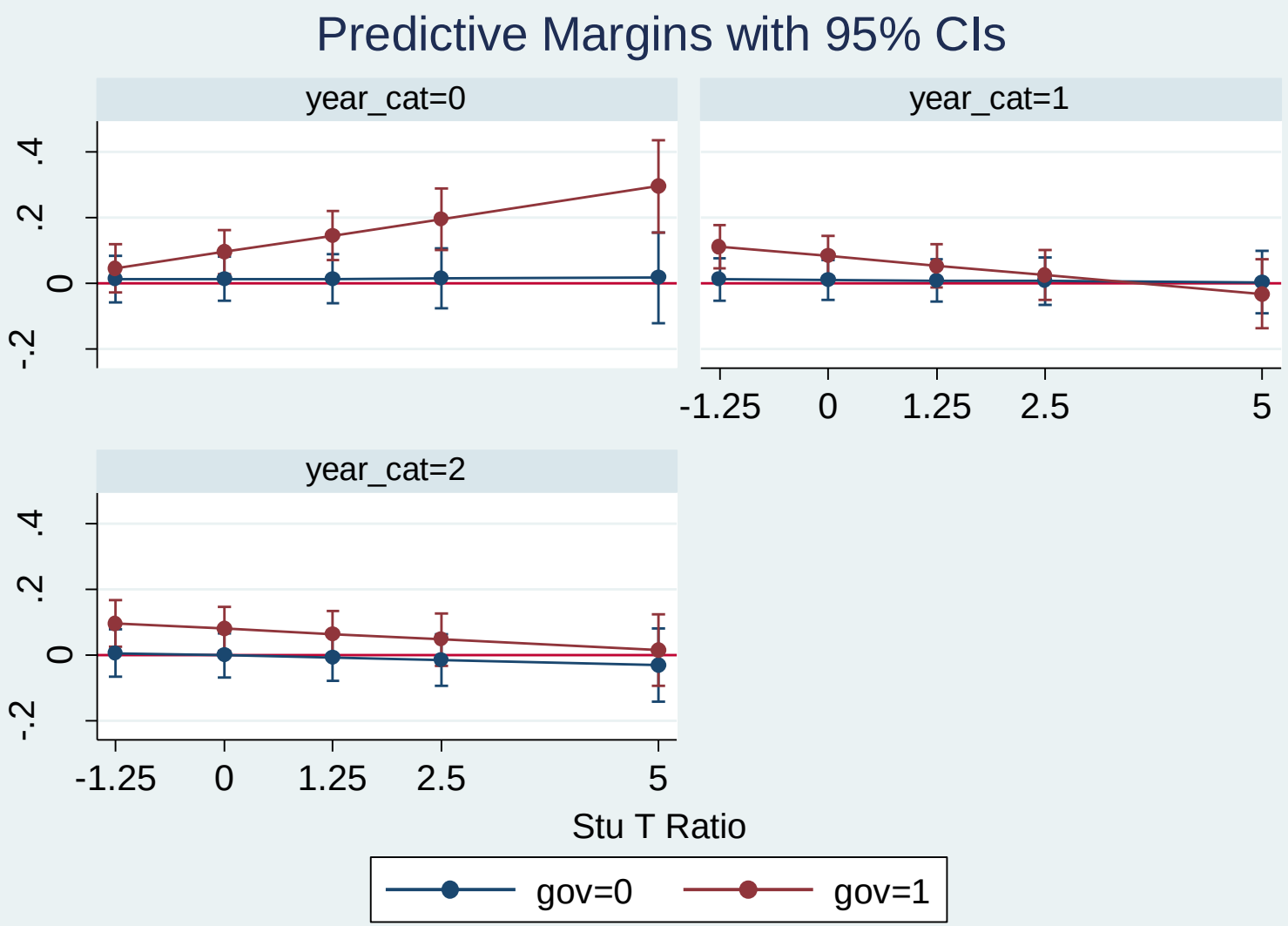
Class size: English



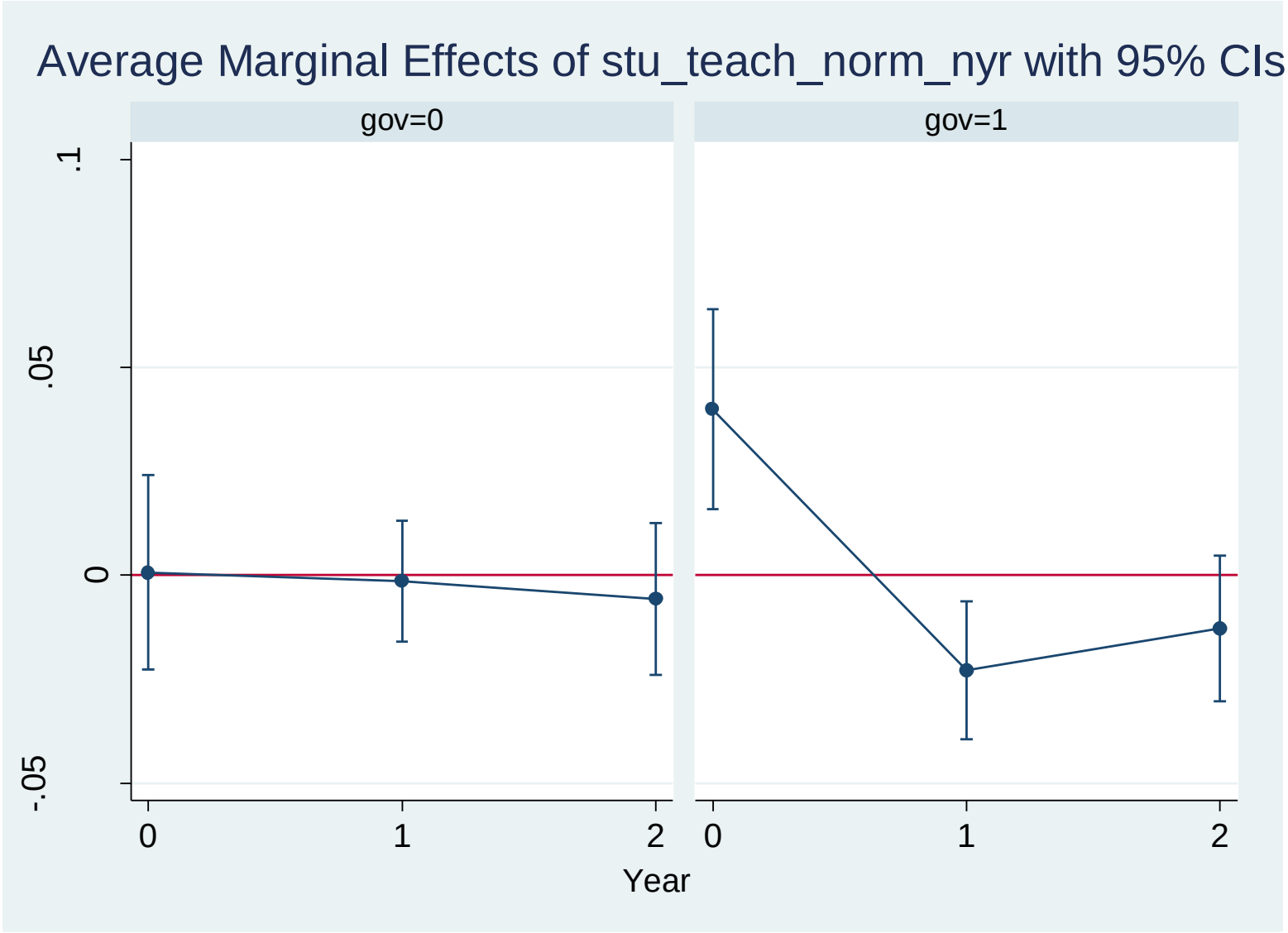
Class size: English



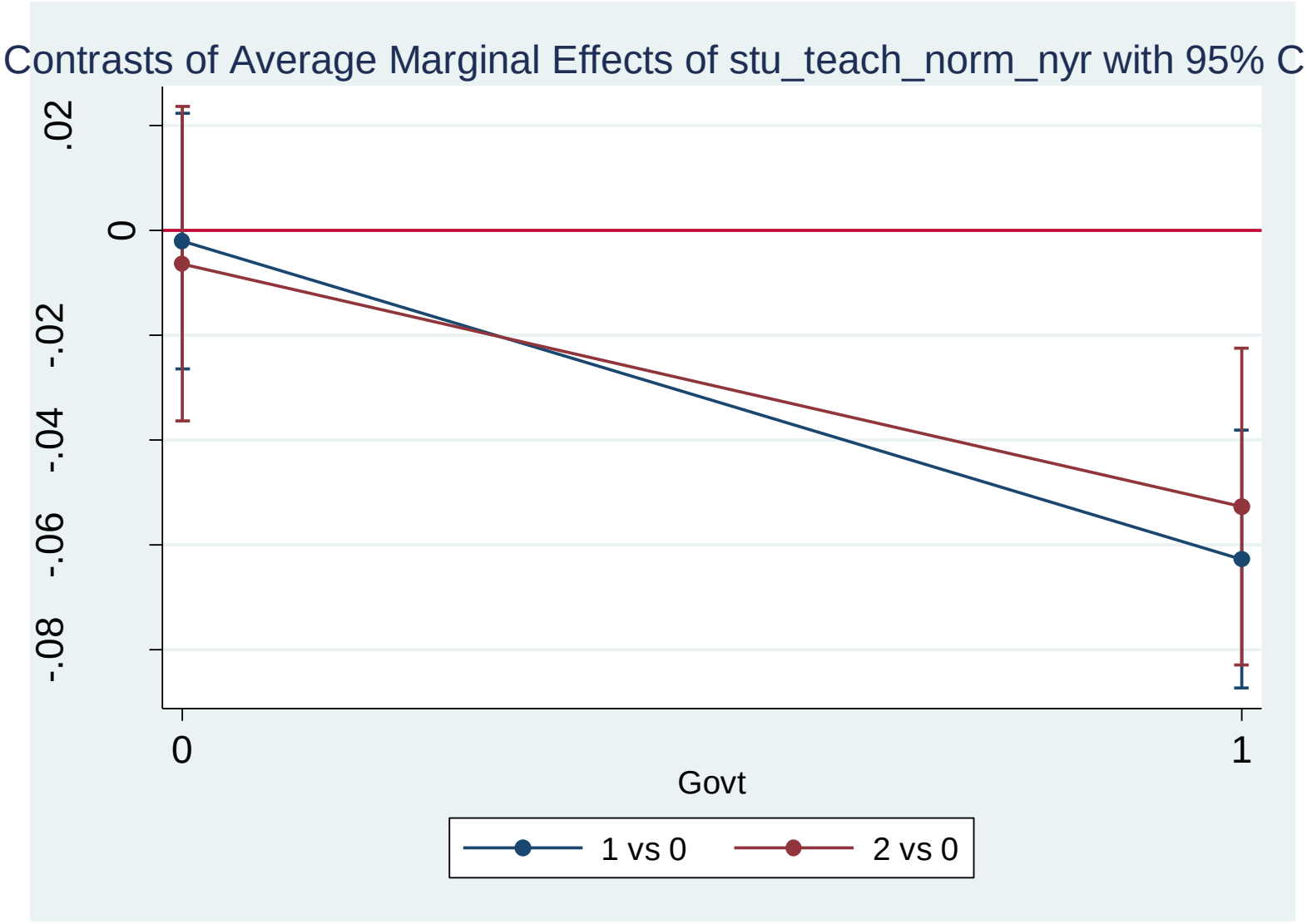
Class size: Maths



Class size: Maths



Class size: Maths



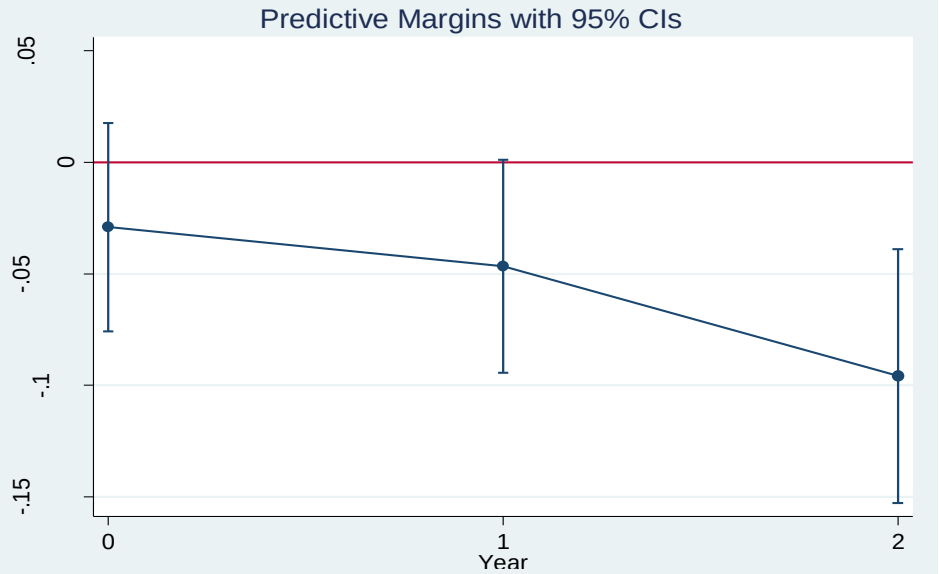
What explains class size effects?

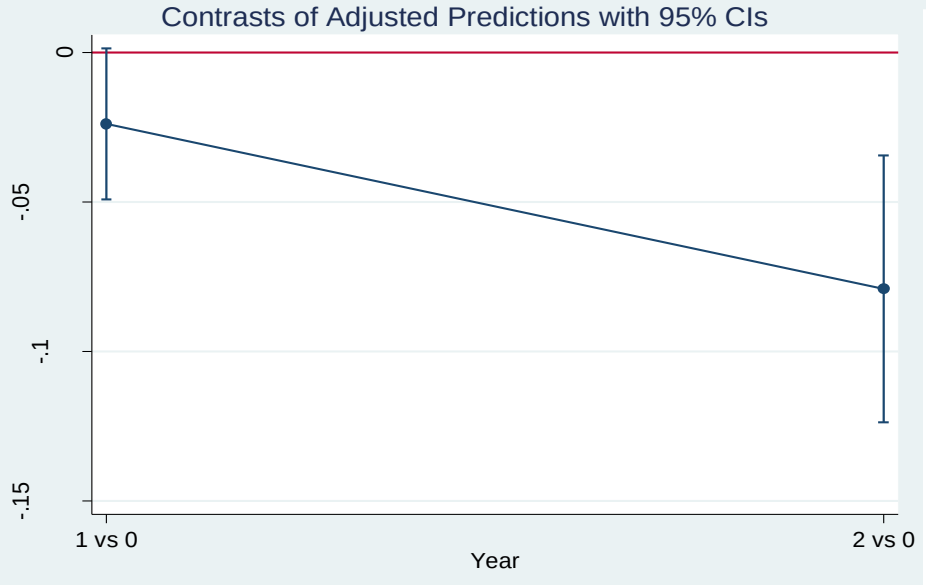
- Class size effect is not necessarily negative.
  - Small classes may not be able to attract school resources other than a teacher.
  - Perhaps low quality teachers are positively associated with small classes.
- However, the increased student numbers in 2012 appear to have negatively impacted learning outcomes.
- The difference between church and govt schools may indicate that institutional effects are important.
  - However, without data on student ability it is not possible to rule out that church schools are attracting lower ability students under the TFF which is really driving the result.

OVERALL CHANGES OVER TIME

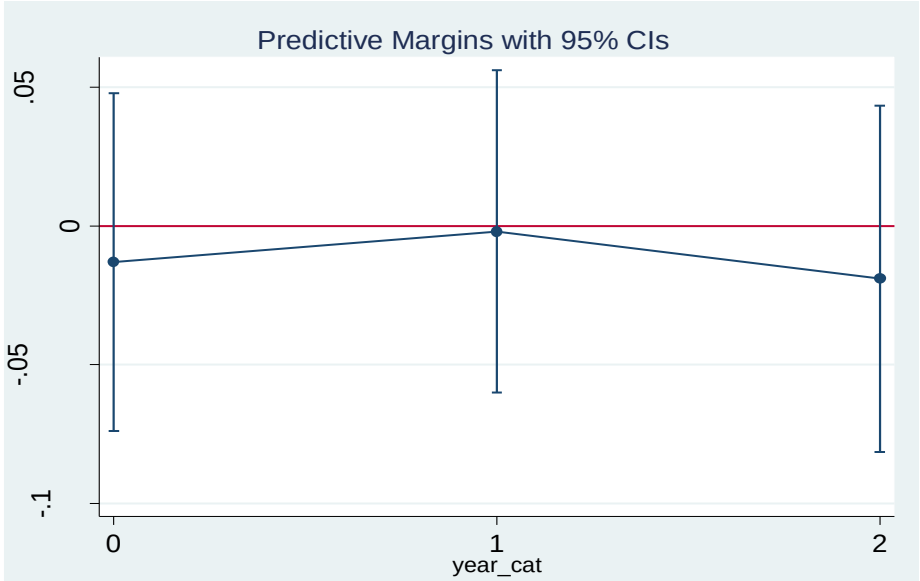
Overall margin- English & Maths

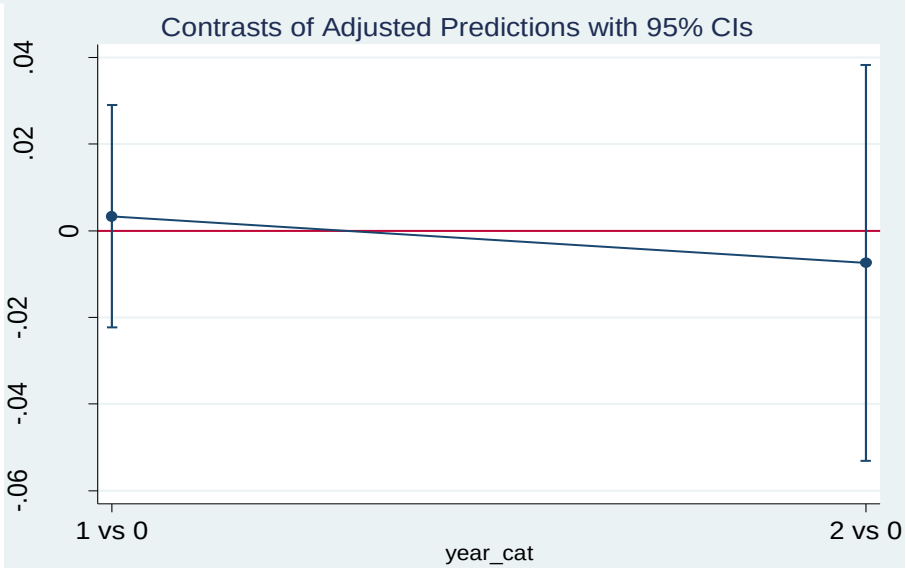
English





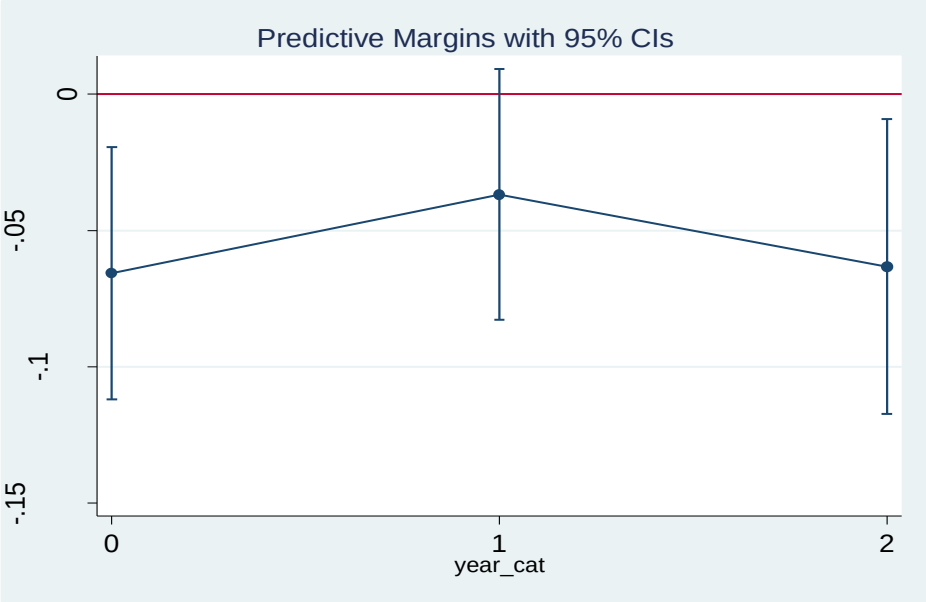
Maths



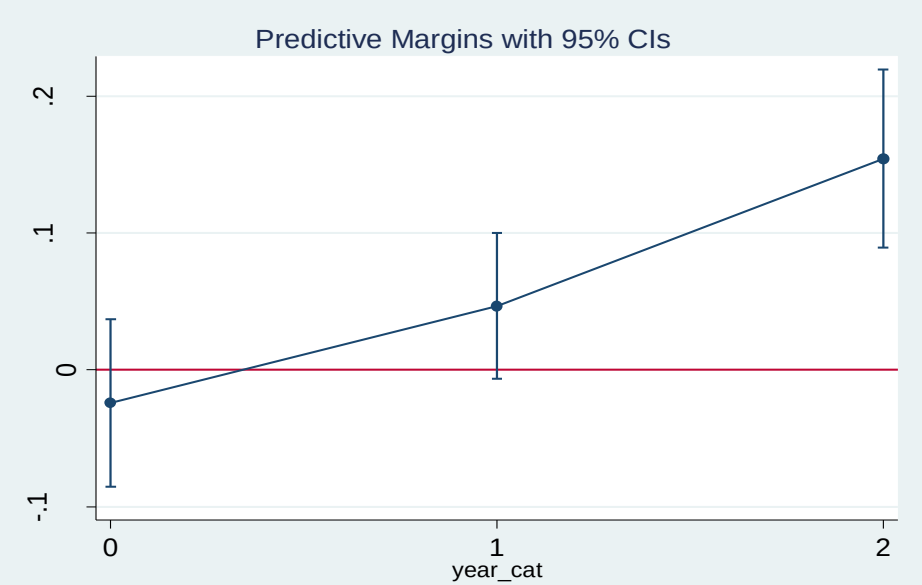


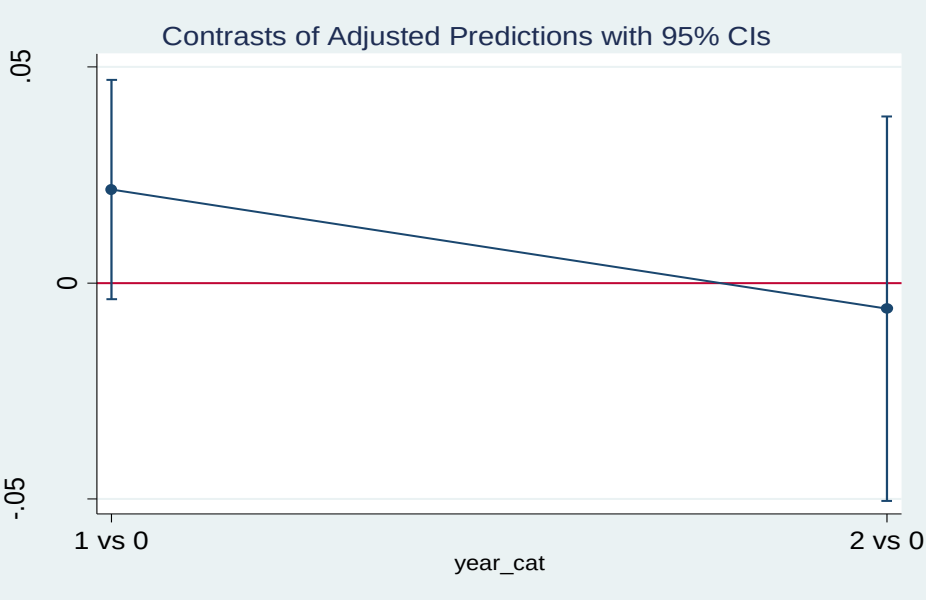
Overall margin- English & Maths

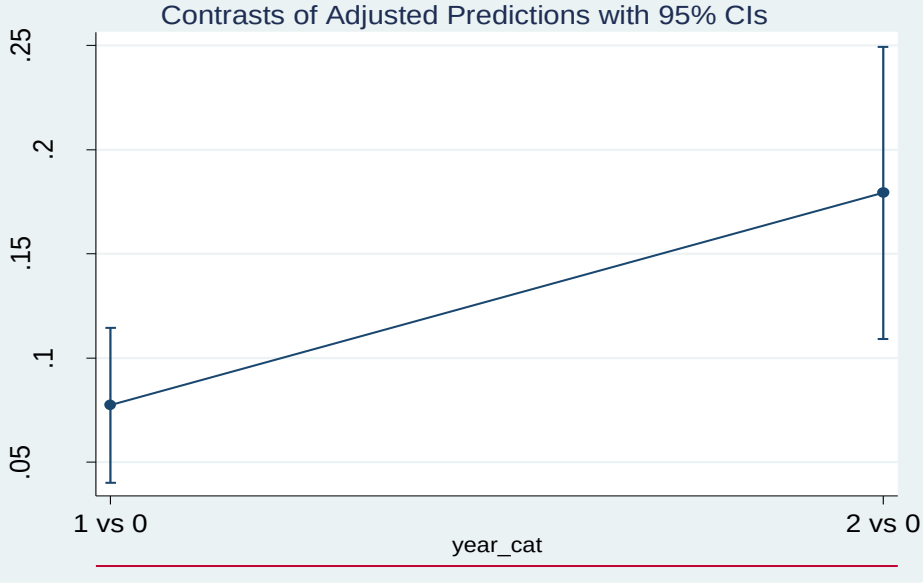
Combined subjects



ELP







POLICY IMPLICATIONS AND CONCLUSION

What reforms might improve education quality?

- Returning / over-age students are disruptive for G8 target age students
  - “second-chance” schooling pathways
  - Community based education
  - After-hours classes
- Double-shift schooling in urban areas to reduce possible negative impact of larger class sizes.
- School based decision making:
  - can lead to large improvements but can be harmful unless coupled with effective monitoring and assessment via centralised external exams.
  - Centralised exams are important – but need to be accessible to families / communities and used to inform what works and what doesn’t.
- School funding adjusted for remoteness.
- Greater number of female teachers
  - Targeting placement of female teachers in areas that have low female student ratios

Conclusion

- Exam results can tell us what seems to matter for learning outcomes.
- However, questions over standardisation over time and location limit what exam results can tell us about the free education policy.
  - Need for controls on standardisation on exams and assessment.
  - The plan to scrap national exams completely is a worry.
- Nonetheless we can conclude that the TFF has had a negative impact on young students i.e. the students who are supposed to be in primary school.
- Big bang approach to education reform can have unintentional effects.
  - Improving access to school is not good enough on its own – cannot assume students already in the system are not negatively impacted by reforms.
  - Why not start reform with a randomised trial or at least have some control group?