Identifying Menstrual Absence Patterns in School Attendance Data

Garima Agarwal, Deepayan Sarkar, Rohini Somanathan

September 14, 2025

ABSTRACT

Despite anecdotal accounts of widespread menstrual absenteeism at school, the phenomenon has not been clearly identified in empirical studies. In this paper, we try to resolve this apparent contradiction by using a novel panel dataset of daily school attendance of a large sample of students from public schools in Delhi. Exploiting the long duration of repeated observations, we employ a generalized linear mixed model approach to identify recurring absenteeism for girls that may be better linked with the menstrual cycle. Despite not finding strong evidence of such patterns we do not rule out their existence. We simulate attendance data with varying shares of menstrual absenteeism to highlight the methodological challenges in identifying menstrual absence without explicit information on individual menstrual timing. Our results indicate that if menstrual absence contributes a small share of total absenteeism, it escapes detection. Findings from primary surveys in the present context support this conjecture. We confirm the general result using another real world dataset of adult workers in a developed country context.

JEL codes: C23, C32, I21, I24

Keywords: panel data, fixed effects, random effects, autocorrelation, education, menstrual absence, school attendance, gender, adolescent girls, menstruation, India

Acknowledgements

The authors thank the Department of Education, Government of NCT of Delhi for granting permission to access attendance data for this study. Data collection was funded by a research grant awarded by the Azim Premji Foundation. Several research assistants and data entry operators, in particular, Pramod Kumar Dubey, have provided critical support in transcribing and cleaning data.

Garima Agarwal: Azim Premji University, Bangalore (garima.agarwal@apu.edu.in); Deepayan Sarkar: Indian Statistical Institute, Delhi; Rohini Somanathan: Delhi School of Economics

1 Introduction

Menstrual hygiene has increasingly received policy attention in the last decade and governments in both developed and developing countries have launched programs to address the specific needs of menstruators.[1] Given public budget constraints, questions are often raised regarding the value of such policies in raising female participation in economic activities like employment and schooling.

In earlier work, Agarwal (2025) shows that a free sanitary napkin distribution program in public schools in Delhi significantly improves attendance for girls in the initial stages of puberty. While the paper argues that the policy reduces unplanned absenteeism, the analysis cannot explicitly show the link with menstrual absenteeism due to data limitations. Therefore, in this paper, we use rich individual-level school attendance data from the same school system to check for menstrual patterns in absenteeism, without explicit information on menstrual timing. We leverage knowledge of the distribution of the parameters of the menstrual cycle in the population alone and apply two methods - autocorrelation functions and generalized linear mixed model regressions. To validate our methods, we apply them to attendance data of adult workers from a developed country setting where no evidence of menstrual absenteeism was found by using a hazard function approach (Ichino and Moretti 2009; Herrmann and Rockoff 2012).

We document an interesting stylized fact about attendance of males and females in both settings and discuss the methodological issues with identifying periodic absenteeism in real world data. There appears to be a high correlation in the daily shares of males and females absent for both adolescents and adults. Further, we find no evidence of periodic absenteeism in school attendance for girls using either correlations or regressions. Despite this, we argue that the existence of menstrual absenteeism cannot be ruled out in our context. We motivate this by simulating attendance data with predetermined menstrual absence patterns and applying the same methods to check for periodicity in absenteeism. We find that these patterns can indeed be detected provided the menstrual signal is strong enough. Therefore, the main results may be on account of a low share of menstrual absenteeism in the school dataset. A primary survey of schoolgirls in public schools in Delhi provides suggestive evidence for this conjecture given that menstruation accounts for 8% of overall absenteeism amongst girls (Agarwal 2022).

Ichino and Moretti (2009), one of the first and few papers in this literature, evaluates menstruation-related work absenteeism as a possible explanation for the gender earning gap for adult workers. Initial findings suggested that women in the reproductive age group (younger than 45 years) have a higher chance of absence spells separated by 28 days compared to similar aged men, which was reported as evidence for menstrual absence in this age group. Herrmann and Rockoff (2012) reanalysed the same data to correct for errors in coding and specification to refute the result. Further, they found a significantly greater chance of such absence spells for older as compared to younger men, neither of whom experience menstruation. No identifiable menstrual absence pattern was found using attendance data of New York City public school teachers either.

The results discussed above suggest that menstruation is not a significant factor in determining work absence for adult women. However, the same may not apply to adolescent girls. Arguably, menstruation could be more disruptive at the initial stages of puberty rather than later in life. This

[1] Countries like India, Kenya, England and Scotland have programs that distribute menstrual products along with some informational campaign (Kalia 2023; BBC News 2017; Scottish Government 2018; UK Government 2023). A small set of countries, most recently Spain, offer menstrual leave as well (Euro News 2022).

is in line with results from Agarwal (2025). It follows then, that if menstruation affects attendance at all, data from pubescent girls would show these patterns.

However, studying menstruation-related absence for adolescents is not straightforward for many reasons. First, menstrual cycles for this age group can range between 20-45 days, which is more than twice as wide compared to the 24-34 day band for adults (Hillard 2008). Second, cycles are less predictable closer to puberty and stabilize with age (Hillard 2002). Third, given school holidays and weekends, we do not observe absenteeism across regular intervals. Moreover, menstrual absenteeism may not occur in subsequent cycles. Fourth, adolescent boys may experience more frequent absenteeism for other reasons, such as pressures of wage employment or general truancy. Therefore, in comparing the two genders, while boys may have a higher overall rate of absenteeism, girls may be expected to display more periodic absenteeism, if menstrual periods matter. Estimation has to allow for individual-specific patterns of menstruation incorporating different menstrual timing in addition to cycle length, as mentioned before. Van Biljon and Burger (2019) use daily attendance data for one academic session for a large sample of students attending public schools in a province of South Africa to study this question by using finite mixture models to estimate menstrual absenteeism. They find a higher probability of menstrual absenteeism amongst young girls closer to puberty as compared to older girls.

This paper adds to the relatively small literature on the effects of menstruation on the economic participation of women and girls in four ways. First, while past studies have focused on employed adults in relatively developed country settings we focus on adolescents from public schools in a developing country where socioeconomic constraints on account of menstruation are likely to be more binding. Further, we use a novel individual panel of daily attendance from a large sample of students as they transition through secondary school. Repeated observations on individuals allow us to estimate recurring absence patterns more accurately. Since individual menstrual cycles have different starting points, aggregate attendance data is not suitable for our purpose since these patterns are lost. Second, from a modelling perspective, this exercise is more challenging for adolescents given their uncertain menstrual status and more variable cycle length compared to adults. We suggest methods to account for these complexities. Third, we suggest improvements in econometric specification over past literature and validate proposed methods using simulated data. Fourth, we attempt to resolve the seeming contradiction between surveys that find high incidence of menstrual absenteeism compared to empirical studies that do not find menstrual patterns in absenteeism. The former studies often ask questions about the last menstrual period, a short recall duration or an overall behaviour (Grant, Lloyd, and Mensch 2013; Vashisht et al. 2018; Krenz and Strulik 2019). Such data does not accurately capture the individual fluctuation in menstrual absenteeism over an extended length of time. Our analysis suggests that if the share of menstrual absenteeism in total absenteeism is low, it may escape detection, especially when information on menstrual timing is limited. Gender-neutral factors in absenteeism also contribute to this. Studies such as Oster and Thornton (2011) that match menstrual timing with school attendance indeed find a low share of menstrual school absenteeism.

The rest of the paper is organized as follows. Section 2 describes the school attendance data used, Section 3 details the econometric methodology for analysis and Section 4 presents results. Section 5 describes the simulation exercise conducted to explain results and Section 6 provides a discussion of findings. Section 7 performs validity checks and Section 8 concludes the paper.

2 Data

Public education systems in developing countries often do not have the funds to invest in technology for record-keeping. This severely limits the scope of research on many important aspects of academic life. Data on attendance are usually maintained as physical records at the school that may not be easy to access or link across years without substantial effort, especially as the time horizon is increased.[2]

Attendance data for this study is drawn from two coeducational schools (henceforth identified by the names "School 1" and "School 2") in the East district of Delhi.[3] Photographs of school attendance registers were transcribed to create a dataset of daily individual attendance. The sample includes all students in grades 6-9 across academic session 2008-14 who remain continuously enrolled in these grades throughout the study period.[4] Further, while an academic session runs from April to March of the next year, this analysis is based on records from July to February given irregularities in records at the beginning and end of the session.[5] This gives us 243 consecutive days in each academic session. The outcome variable indicates whether a student is absent on a particular day, treating all other days when they are present or official school holidays as the base category.[6] [7]

Table 1: Number of students whose attendance was recorded in the study, by school and gender, along with the overall rate of absenteeism (percent) within each group. Sundays are excluded from the calculations, but for holidays, all students are assumed to be present.

gender	school_id	Students	Overall absent %
female	1	797	12
male	1	1108	17
female	2	759	8
male	2	1100	9

[2]A few studies have looked into the reliability of school attendance records compared to random spot checks to find conflicting results. While Benshaul-Tolonen et al. (2019) find missing official records on days of researcher-reported absenteeism, Baird, McIntosh, and Özler (2011) find near perfect matches between the two sources. The validity of school records may indeed be context dependent and merit further study, but spot checks would be infeasible for studies with a long time horizon such as the present one. Moreover, for this dataset, it would be hard to argue that school records were differentially biased for the genders or that certain kinds of absences were more likely to be be mis-recorded.

[3]One school had grades 6–12 and the other had grades 6–10 for the entire study period. Primary grades were added in the latter towards 2013.

[4]If students stay in the same school across the study period they can be tracked as they transition across grades. We exclude students who drop out mid-session or leave the school prematurely during the study period. However, there are cohorts we observe for less than four years. For example, those in grade 9 in 2008 are only observed for one year.

[5]At the start of the session in April, students are promoted across grades and transferred across sections making it difficult to track attendance. The following months of May and June are largely covered by the summer vacation. Finally, examinations are held in March, with alternate days designated for preparation and tests, making it difficult to study attendance patterns.

[6]As mentioned in (Agarwal 2025), daily attendance is recorded as 'present', 'absent', 'on leave' or 'name struck off'. The last category is for students who are removed from the rolls mid-session, often due to long absences. These students are excluded from the sample. Entries of 'absent' indicate that the student did not submit a prior application for the holiday whereas, 'on leave' is for a holiday taken with an application. Holidays with prior application are more likely to be planned and therefore, will not contribute to the exercise of identifying menstrual absence. For analysis, only 'absent' days are considered and 'on leave' days are coded as 'present'.

[7]Sundays are treated as missing.

Table 1 gives the number of students and rate of absenteeism in the sample by school and gender. Data for a total of 3,764 students were recorded - 1,556 girls and 2,208 boys.

Following Agarwal (2025), Figure 1 describes the age distribution of the girls in the present sample along with lines denoting the upper (13.76 years) and lower (12.4 years) limits of mean age at menarche as recorded in the literature (Khadgawat et al. 2016; Pathak, Tripathi, and Subramanian 2014). It seems reasonable to expect that a majority of girls in grades 7-9 in the sample would have crossed puberty while the share for grade 6 appears lower. The dataset does not contain information on individual menstrual status, timing of period, cycle length or reasons for absenteeism.

Overall, the rate of absenteeism is around 10% and a total of 258,960 absent days are recorded in the sample. These work out to 178,227 absence episodes counting spells of absences spanning consecutive days only once. On average, a student takes 25 days off in a year. Absenteeism is lower among girls than boys with 21 and 28 total days of absence, respectively. The average number of days taken off in a month are 3 and 4 days for girls and boys, respectively. While absence episodes range from 1-6 days, most are 1 or 2 days long.[8] In fact, the gender distribution of these spells is almost identical with about 73% single-day and 17% two-day spells. The remaining 10% spells are 3-6 days long. Figure 2 plots the distribution of absence spells for both genders.

By disaggregating the rates of absenteeism, we get a more nuanced picture. Figure 3 plots the overall rates of absenteeism per day for each gender group separately for each school.

Another interesting feature of these attendance records is the strong correlation between the share of girls and boys absent on a given day (Figure 4). The same pattern holds across both schools. While this correlation is partially explained by the trend of increasing absenteeism over time seen in Figure 3 and day-of-the-week effects, there appear to be additional gender-neutral factors affecting absenteeism.

Figure 5 similarly shows that rates of absenteeism vary widely across students. These significant sources of variation (by day and by student) make it potentially challenging to detect menstrual absenteeism patterns in the data.

3 Methodology

Our broad goal is to identify periodic signals in the individual time series of absence data, while accounting for the other systematic sources of variation present in the data. In a time series data with no additional sources of variation, sample autocorrelation functions (ACF) can reflect periodic signals in the data. We start with this, but unfortunately, given the various other sources of variation in this data, a simple ACF is unlikely to be helpful. Our preferred method is a generalized linear mixed model (GLMM) to account for background sources of variation, and we use the essential idea of autocorrelation to try and detect periodic effects.

[8] Since the Kishori Scheme was in operation from December 2011, the sample was split into before and after periods for both genders to check if overall absenteeism or length of absence spells reduced after program implementation. This does not appear to be the case.

3.1 Autocorrelation Functions (ACF)

Autocorrelation functions (ACF) provide a way to quantify the strength of the relationship between two values of a variable measured at different points in time as the separation between them changes. A random process should have zero correlation at all lags, while a process with a periodic component would have non-zero correlation values at least at some lags. Evidence for menstruation-related periodicity may be expected to show up as spikes in the ACF in the neighbourhood of lag28, i.e. day (t - 28), for females and not males.

Values of the ACF at each lag are first calculated for all individuals and then an average value is considered for each lag for girls and boys, separately. Raw absence data is adjusted by differencing out day and individual-specific absence probabilities to assess correlation between absences driven by factors other than these.

3.2 Generalized Linear Mixed Model (GLMM)

Generalized linear models are a natural choice for problems such as the present one which involve modelling a binary outcome variable conditional on covariates. A mixed effects model further allows for the inclusion of both fixed and random effects. We first fit a model with separate (fixed) means for each school and gender combination, an additive fixed effect for day-of-the-week, along with further random effects for each day and student. We fit both a standard linear model (which assumes equal error variance for all observations), as well as a logistic regression model which more appropriately models the response as binary outcomes, with the log-odds of absence probability linearly depending on covariates. Although the logistic model is more appropriate, it is computationally more challenging to fit, both in terms of speed and accuracy.

We hope that this model substantially accounts for most of the systematic patterns seen in our earlier exploratory analysis. The model still assumes that the random effects for students are uncorrelated, which may not be true. However, we ignore this issue for now as modeling it is too challenging, and we expect the effect to be small. Our hope is that if there are any periodic signals in the data, they would be easier to see in the residuals from this model.

We consider the residuals from this model and compute something similar to the ACF. However, we need to be careful as the dates are not regular since we do not observe March-June. For each value of lag k from 1 to 50, we compute the lag-k autocovariance as follows: first we split the data by each student and session, and consider the residuals of the 243 consecutive days in this session. The lag-k autocovariance is the mean of the lag-k crossproduct of the residuals (excluding Sundays). These means are then averaged across all student-session combinations to estimate an overall lag-k autocovariance. The resulting values, as a function of the lag k are shown for males and females in Figure 6. There do not seem to be any systematic differences between the functions for males and females, as we might have expected to see had there been extra periodic signals in the absenteeism pattern for females.

3.3 GLMM incorporating lags

A more formal procedure to incorporate lags into the model is to add them as fixed effects. Due to speed an convergence issues with logistic mixed effects models, we henceforth consider standard linear models instead of logistic models, even though the response is binary.

The specification includes random intercepts for individual and date indicators as well as fixed effects for lag1, lag7 and lagX where X $\in$ [8,45]. Absence on consecutive days is captured by including lag1 and weekly absence patterns are accounted for by including lag7. Further, lag8-lag45 are sequentially included, one at a time, to check if any of these are strong predictors of absence, controlling for the effects of lag1 and lag7. In previous analyses there has been a focus on controlling for lags that are multiples of seven only. This specification is flexible to other kinds of periodicity as well.

The specification described above takes the following form for individual i at time t, incorporating lagX:

$$Y_{it} = (\mu + \alpha_i + \gamma_t) + \beta_1 lag1_{it} + \beta_7 lag7_{it} + \beta_X lagX_{it} + \eta_s + \delta_t + \epsilon_{it} \tag{1}$$

The dependent variable Y_{it} is an indicator of absence for individual i at day t. Coefficients α_i and γ_t represent random intercepts for individual and date effects, respectively, while μ is the common intercept. β coefficients capture the fixed effect of the respective lagged values of absence in predicting Y_{it}. We also incorporate fixed effects for the school (η_s) and day of the week (δ_t).

An improvement over previous literature, through either ACFs or mixed model regressions is in the inclusion of date-specific effects which may be important given the strong observed correlation between daily absences for both genders.

4 Results

Figure 7 presents the ACF plot for school attendance data after adjusting the absenteeism variable for day and individual-specific absence probability. A red line denotes the position of lag28. There appears to be a downward trend in correlation with higher lags and no prominent spikes emerge around lag28 for girls.

Figure 8 presents results from the linear mixed effects model specification. Each bar represents the coefficient estimate for a particular lag (X $\in$ [8,45]) included one at a time in addition to the base model of lag1, lag7 and random intercept terms for individual and date-specific effects. The graph shows no spikes for lags around 28 (denoted by a red line) for girls and a similar downward trend in coefficients for higher lags is visible.

Neither method finds strong gender differences in absenteeism. We specifically check for patterns in the period prior to the sanitary napkin distribution program studied in Agarwal (2025), and results remain similar, even on disaggregating by grade.

5 Simulated attendance data

To assess whether the proposed methods are appropriate for the empirical exercise of the previous section, they are validated using datasets with pre-determined absence patterns. This helps assess the ability of the methods to detect menstrual absence of varying degree in a controlled setting. The following describes the procedure followed to create these datasets.

Suppose absences for each student are on account of three factors - systematic, random or menstrual period-related. Systematic factors include day and individual specific absence probabilities for which priors can be drawn from actual data. These allow for the possibility that some students are more prone to absenteeism than others and some days are high absence days for all, perhaps due to weather-related disruptions or festivals, for instance. The random component incorporates unpredictable absences. Lastly, menstrual period absence is applicable only to females. Period cycle length is modelled normally distributed with mean length of 28 days and a standard deviation of 2 days. One of the first 4 days of each cycle is randomly assigned as a candidate day for absence.

The total absences in the dataset (based on actual data), are split across these three components in predetermined proportions. For each student on each day, potential absence is predicted for every component based on Bernoulli draws. The success probability for each of these Bernoulli draws depends on individual absence rate and in the case of systematic factors, also on day-specific absence rate. Finally, a student is observed to be absent on a given day if any of the three components gets assigned as a potential absence for that day.

Using priors based on grade 9 students in 2011 from one school in the dataset, multiple datasets are simulated by altering the share of the period component as 15, 20 and 30%. The total number of students modelled is 3,200 - 1,280 girls and 1,920 boys, to reflect the actual gender composition from this data. The total number of absent days in simulated datasets are between 81,000-84,000. Proposed methods should detect a stronger menstrual absence pattern in datasets with a higher share of period absences for girls.

Figure 9 presents ACFs for simulated datasets for girls and boys separately. For girls, a higher correlation with lags around 28 starts to become apparent with increased period share. This is especially clear in the bottom panel of the figure for the dataset with 30% period share. Not much changes for boys across these graphs. Figure 10 shows coefficient estimates for the included lag term in a linear mixed effects model described in specification Equation 1, as period share changes from 15 to 30%. Estimates are positive for girls in the neighbourhood of lag28, even for 15% period share and gradually become more pronounced as period share grows. The same is not true for boys.

Both sets of graphs suggest that if the period signal is strong, these methods are able to detect it.

6 Discussion

As mentioned before, menstrual patterns in absence are hard to detect for both adults and more so for adolescents. For the latter, the process of identification is further complicated on account of unstable menstrual cycles at this life stage. In data from schools in Delhi, analysis is constrained by a lack of information on reasons for absenteeism, menstrual status, individual menstrual timing and cycle length. Further, plots of daily absence shares across girls and boys show a high correlation, indicating the presence of strong gender-neutral factors in determining absenteeism.

In this analysis, two methods are proposed to check for the presence of menstrual absence patterns - autocorrelation functions and generalized linear mixed models. Neither method shows any evidence for menstrual absence even after controlling for individual and date-specific factors. While these findings are in line with the literature for adult work absenteeism (Herrmann and Rockoff 2012), it may not be straightforward to extend the same conclusion to school data for adolescents.

Analyzing results from a series of simulated datasets, it seems that both methods can detect menstrual absence if its share is large enough. In particular, if menstrual absence accounts for 20-30% of the total absence in the dataset, then the ACF begins to show higher correlations in the neighbourhood of lag28, while no discernible pattern is visible if period absence share is 15%.

Regression models show that coefficient estimates in the neighbourhood of lag28 become more apparent with greater period share. Therefore, it remains possible that the lack of evidence from actual school data is on account of a low share of menstrual absence rather than the non-existence of this phenomenon. A primary survey of schoolgirls in public schools in Delhi, in fact, finds that the share of menstrual absenteeism is 8% for the sample (Agarwal 2022). Further, among the girls who report such absenteeism, it accounts for almost 30% of overall absenteeism. This indicates, that there maybe subpopulations that are more affected by menstrual constraints within a school body.

Essentially, if menstrual absence is not a ubiquitous phenomenon, then identifying such patterns in absence may not be simple given individual variation in period timing and cycle length and the presence of other strong signals such as gender-neutral day-specific absenteeism. The dataset used in Oster and Thornton (2011) does collect information on menstrual timing and absence but from a small sample over a relatively short period while an intervention was running. Further, there seem to be issues of missing data as well. Nevertheless, 50% of surveyed girls report some menstrual absenteeism, but it is around 14% of the total absenteeism reported over the school year.[9] Given the limited empirical knowledge regarding menstrual absenteeism, detailed data collection in diverse settings is needed. Many other studies record the incidence of period absenteeism at the extensive margin, but do not collect detailed information regarding the extent of such absenteeism over an extended duration. These statistics vary across settings and are based on whether or not an individual experienced any absenteeism in the last menstrual cycle, or over a short recall period. For instance Grant, Lloyd, and Mensch (2013) find that 32% period absenteeism among schoolgirls in rural Malawi, Vashisht et al. (2018) report 41% incidence among schoolgirls in urban Delhi and Krenz and Strulik (2019) capture 15-20% incidence among working women in Burkina Faso.

In the context of education policy and menstrual absenteeism, this finding is useful if limited program funds have to be rationed across regions but the extent of menstrual absence is unknown, as is mostly the case. It may be simpler to evaluate patterns in attendance records rather than collect individual data on reasons for absenteeism. The methodology proposed here would then help identify regions with greater menstrual absence based on attendance records alone. In fact, this methodology may have applications in other fields where a cyclical phenomenon that is hard to observe is linked with another event that may be simpler to record and evaluate for patterns.

7 Validity checks

An additional check on the validity of proposed methods comes from applying them to attendance data in a different context. The dataset used by (Ichino and Moretti 2009) captures daily attendance for all employees of a large Italian bank who were on payroll continuously from 1993 to 1995. Only illness-related absences are considered here, a marked difference from Delhi school data where all absences are recorded. Of the 14,857 workers,[10] 2,965 are women and 11,892 are men. On average, 10 days of absence are taken in a year. In contrast to school data, absenteeism is higher among female employees. While women take about 13 days in a year, men take about 9. The median shares of daily absenteeism are around 4 and 2% for women and men, respectively. The upper limit of absenteeism is around 9 and 6% for women and men, respectively, and is much lower compared

[9] In this context, out of the 180 day school year, girls report attending 86% i.e. overall absenteeism is around 25 days. Further, around 3.5 days of absence are due to periods, which gives us a share of around 14%.

[10] Amongst the 16,208 workers, those with at least one illness-related absence are considered for the final sample.

to school data. The level difference is plausible given that work absenteeism due to illness amongst adults in a developed country is likely to be lower than school absenteeism due to all reasons amongst adolescents in a developing country.

Interestingly, Figure 11 shows that daily absence shares for both genders are strongly correlated for this dataset just like for school attendance (Figure 4). In fact, applying the two methods of the ACF and GLMM regressions to this data and comparing younger and older subsamples of women gives qualitatively similar results to the main analysis. Figure 12 and 13 show no evidence for menstrual absence, in line with the conclusions from (Herrmann and Rockoff 2012).

8 Conclusion

This paper explores the phenomenon of menstrual absence among adolescents by studying individual attendance data from coeducational public schools in Delhi. While the literature has found no evidence for menstrual absence among adult women, this issue is relatively understudied in the context of adolescents, who may, in fact, face greater constraints in dealing with menstruation for the first time.

Proposed methods do not find evidence for menstrual absence in school attendance data. However, simulating a series of datasets with pre-determined menstrual absence shares helps build a clearer understanding of data requirements in order to successfully identify a menstrual absence signal. It seems that methods are able to detect menstrual absence if it accounts for a share of about 20-30% of overall absence in the dataset. The lack of menstrual patterns in our data could then be explained by a low share of menstrual absence in overall absenteeism. A small body of existing literature supports this empirical possibility. Given challenges of data collection, it may be that the available datasets are drawn from settings where menstrual constraints are less binding and thus, menstrual absenteeism is lower.

While providing a way to reconcile anecdotal evidence about menstrual absence with a lack of empirical evidence from various contexts, this paper also highlights the need for data collection in diverse settings to generate estimates of the share of population affected by menstrual absenteeism and the size of the effect. This data could be useful in taking a holistic look at policies encouraging school participation for adolescents.For instance, (Van Biljon and Burger 2019) estimate greater menstrual absenteeism among girls from poorer schools in South Africa. It would be interesting to evaluate these patterns for a rural Indian setting, where menstrual absenteeism may be higher than in the dataset from Delhi.

Our methodology may have useful implications in identifying populations with greater extent of period absenteeism for better program placement. Further, there may be analogous phenomena in other fields where patterns in an observable event may help indicate the extent of an unobservable signal of interest and feed into decisions about policy.

9 Figures

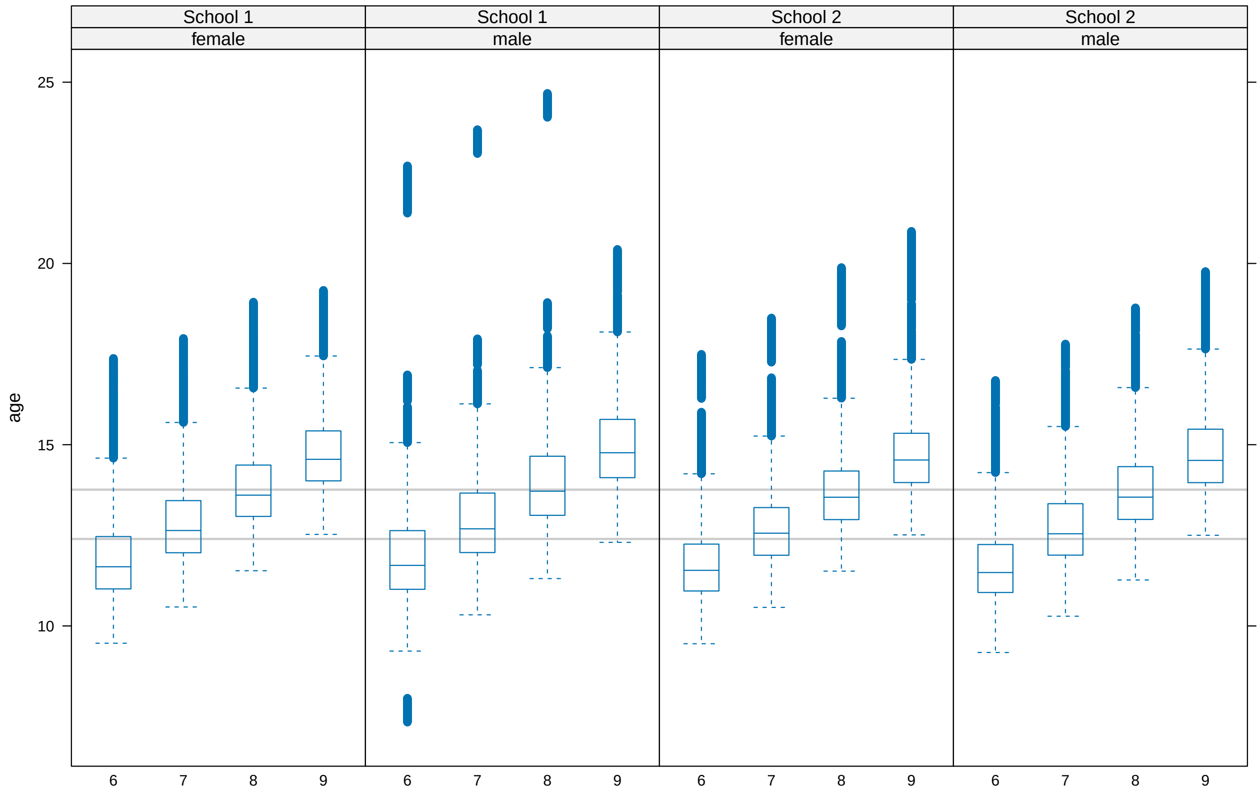


Figure 1: Age distribution of girls (2008-14)

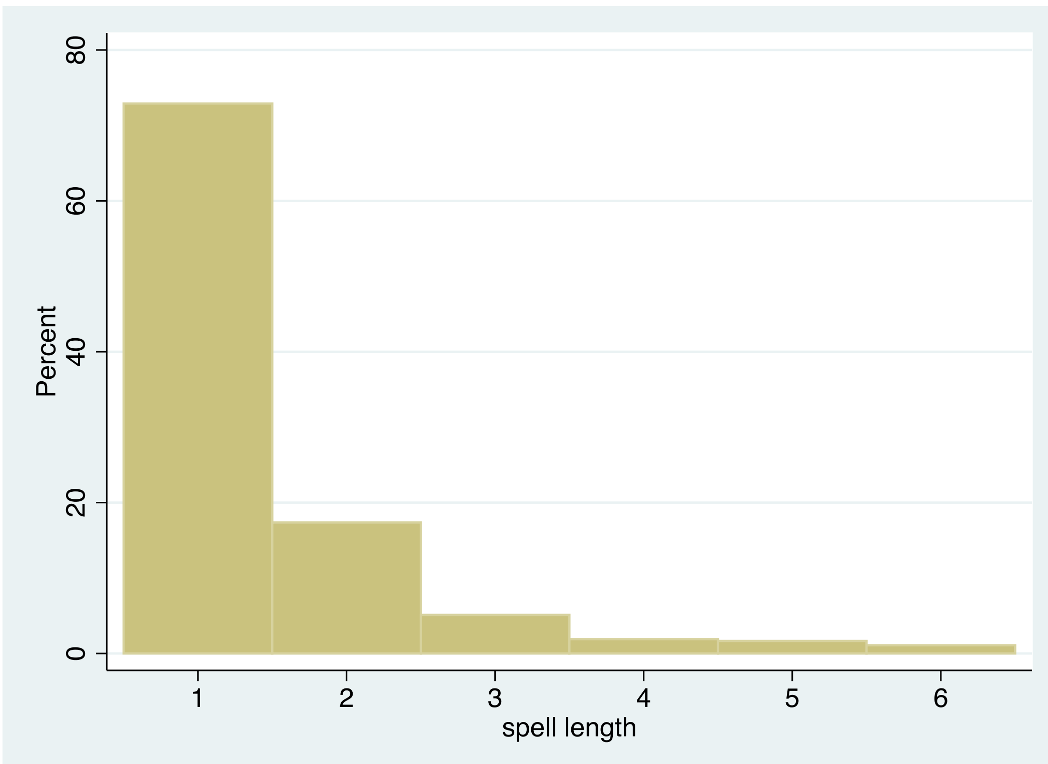
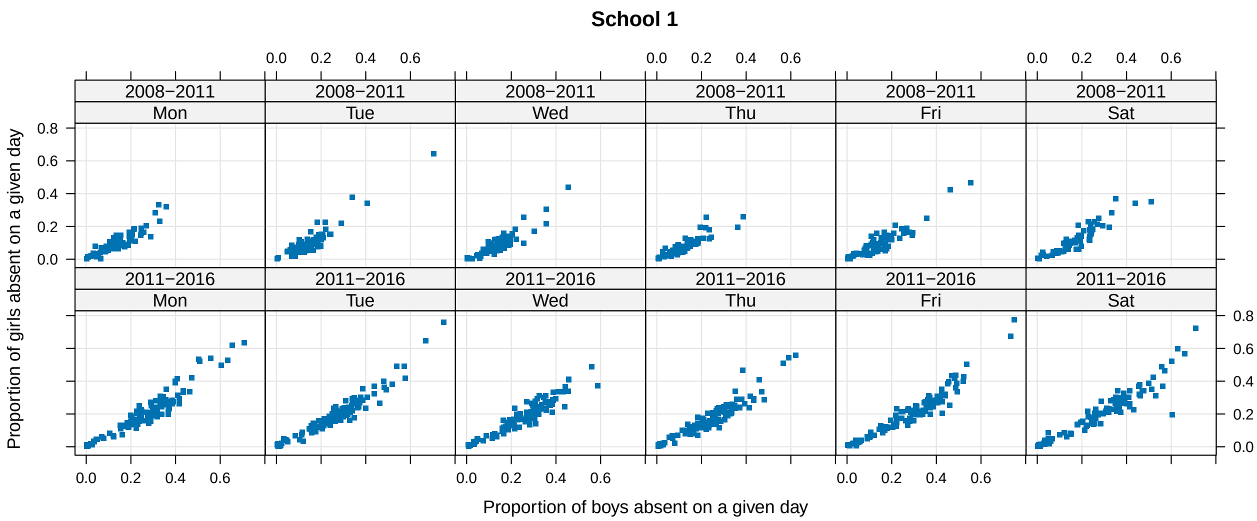


Figure 2: Distributions of absence spells (2008-14)

Figure 3: Proportion of students who are absent on any particular date over the course of the study. Sundays and holidays are excluded. Rates of absenteeism are generally lower among girls vs boys and in School 2 vs School 1, as would be expected from Table 1. However, there is also a distinct increase in the overall rates of absenteeism among all groups, starting from 2012.



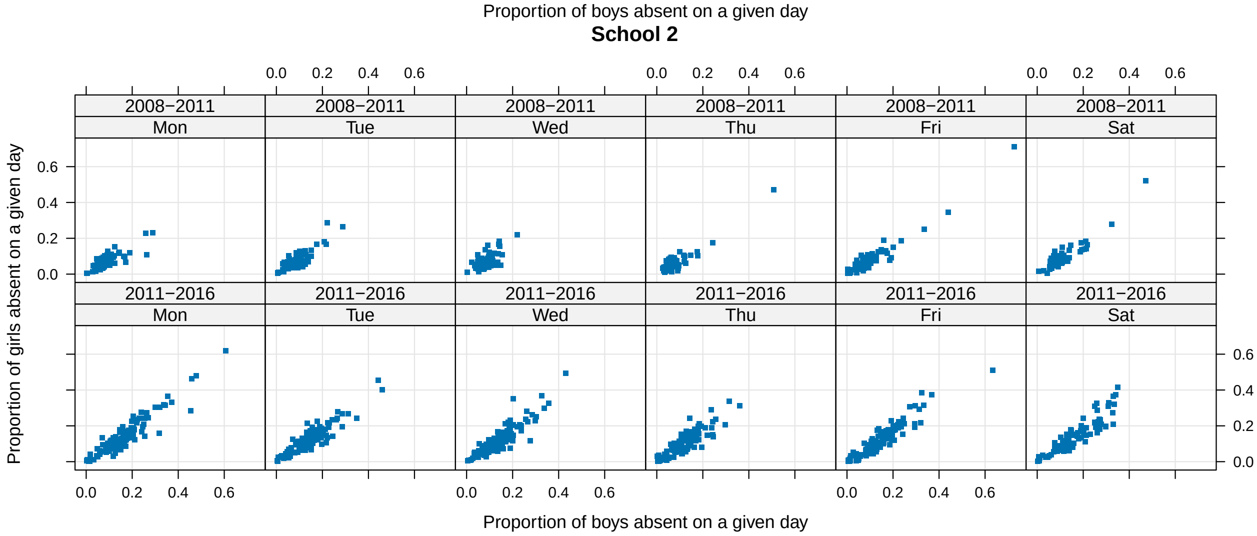


Figure 4: Correlation between daily absenteeism rates in males and females, before and after mid-2011. The data are segregated by day-of-the-week, but this explains the correlation only partially.

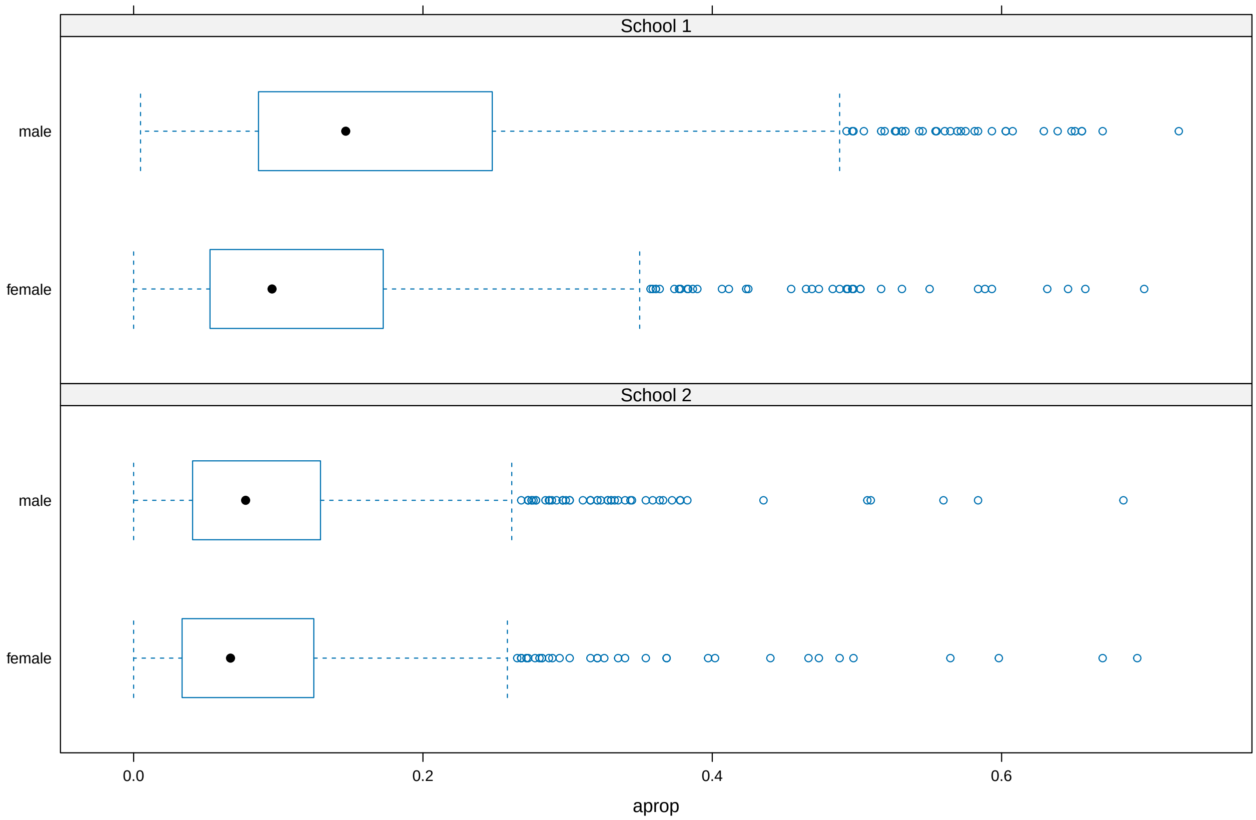


Figure 5: Comparative box-and-whisker plot showing distribution of the proportion of absent days by student. The distribution is heavily skewed (median rates of absenteeism are significantly lower than the average rates given in Table 1), but many students have very high rates of absenteeism, especially in IP.

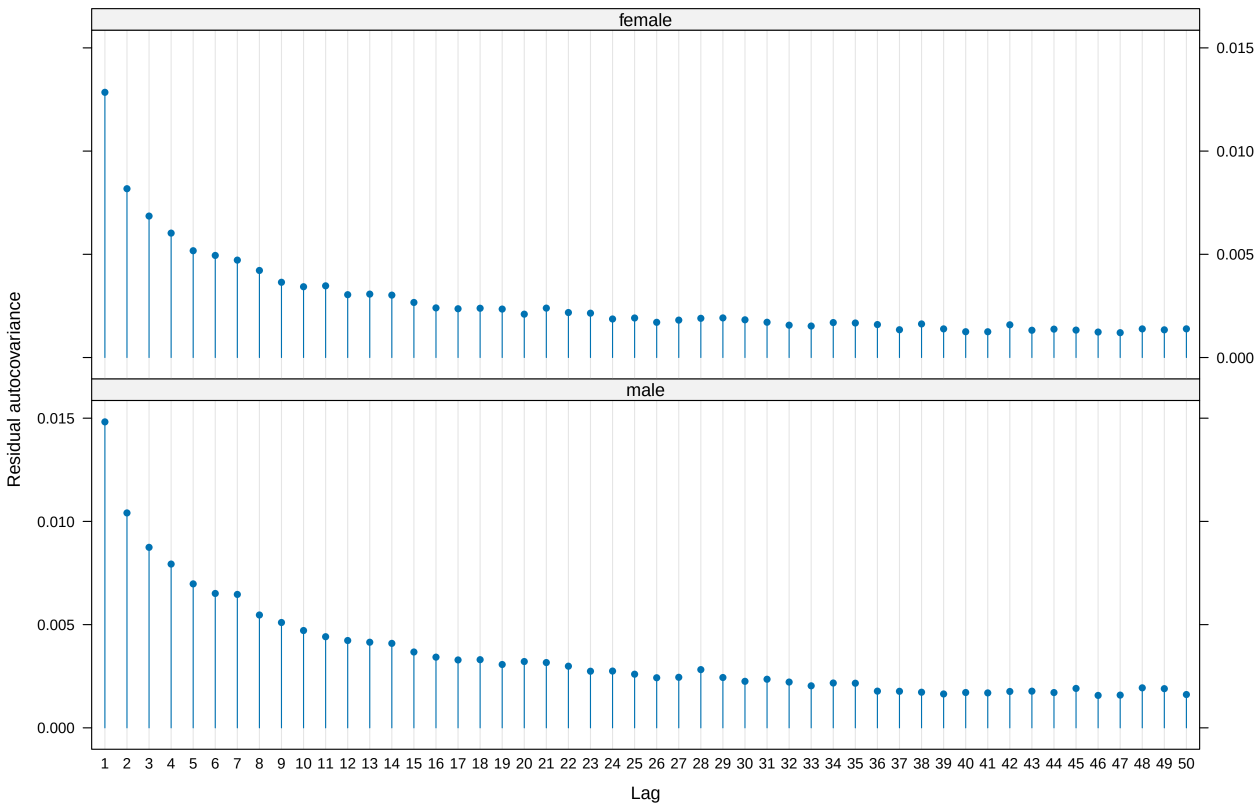


Figure 6: Autocovariance estimated from the residuals obtained after fitting the mixed effect logistic model. We would expect periodic components of lag around 30 to show up as a spike around that lag. However, no such lag is apparent, and there do not seem to be any noticeable differences in the autocovariance pattern for males and females.

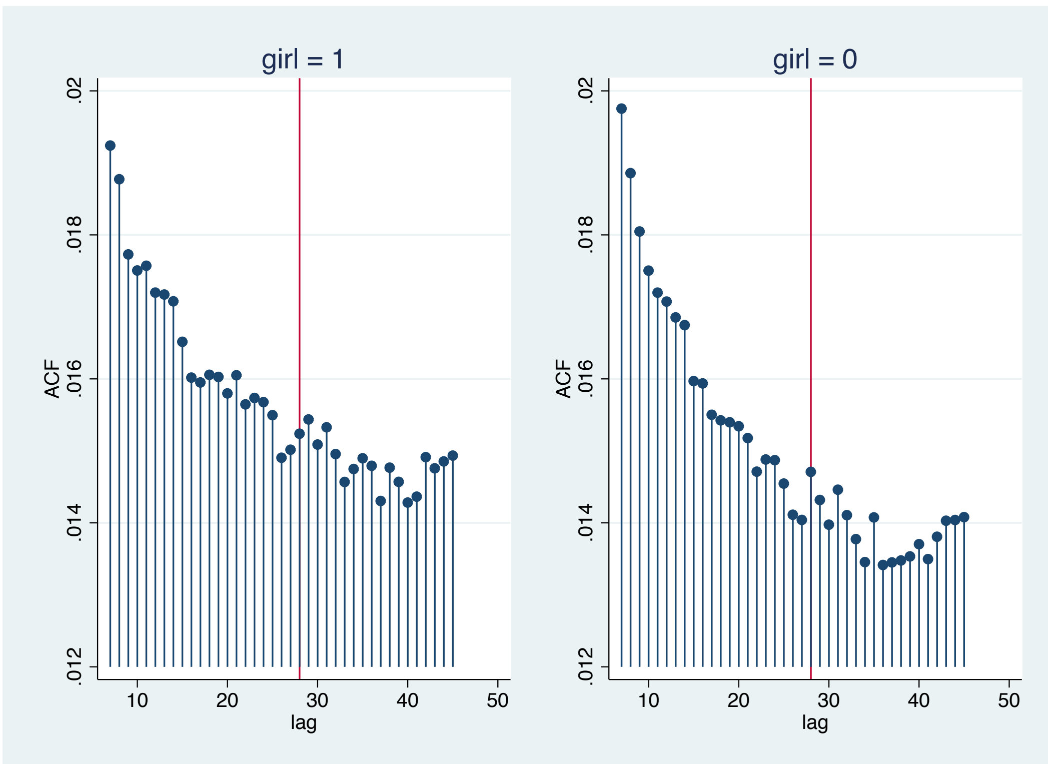


Figure 7: Aggregated autocorrelation function plot for school data (2008-14)

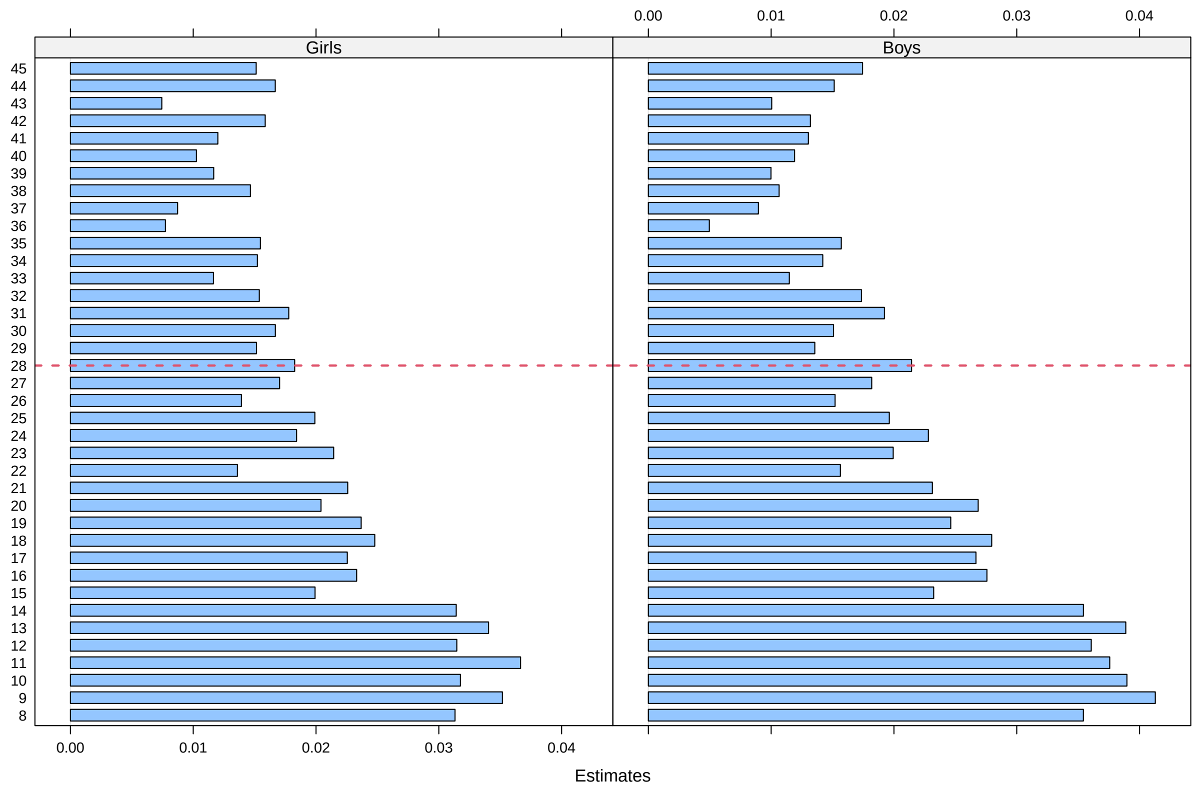
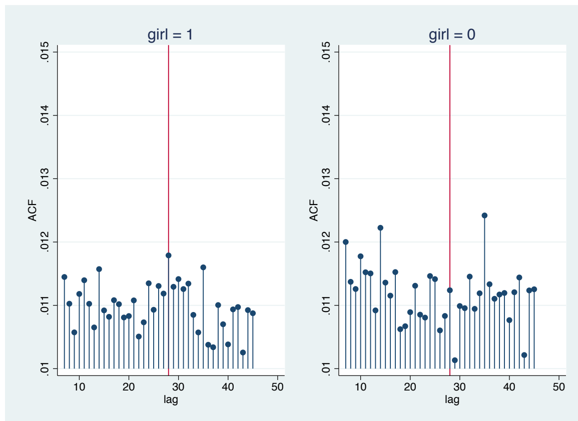


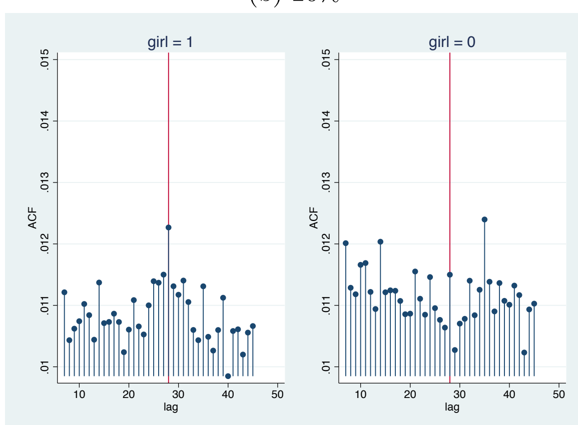
Figure 8: Regression coefficients for $lagX$ for school data (2008-14)

Figure 9: Aggregated autocorrelation function plot for simulated data by period share

(a) 15%



(b) 20%



(c) 30%

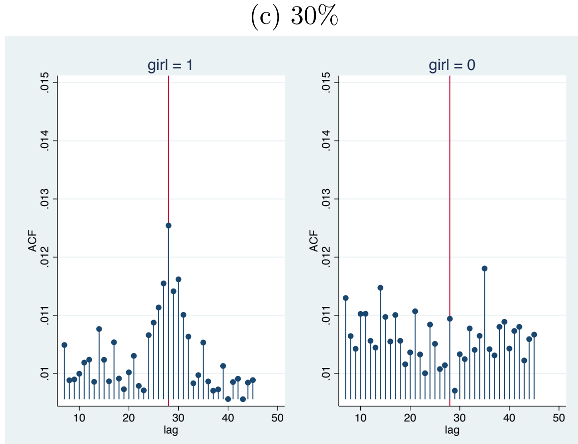
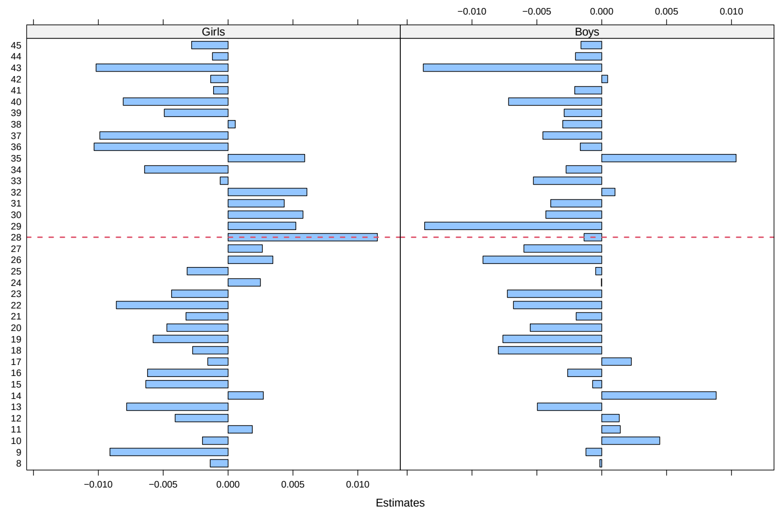
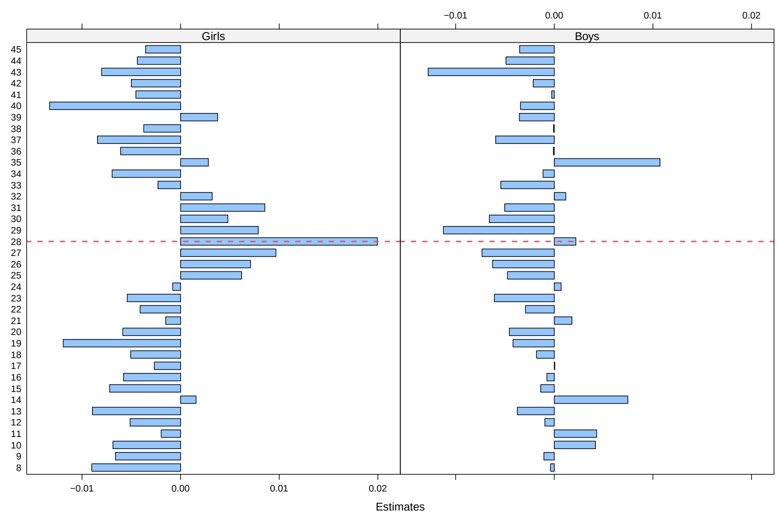


Figure 10: Regression coefficients for $lagX$ from simulated data by period share

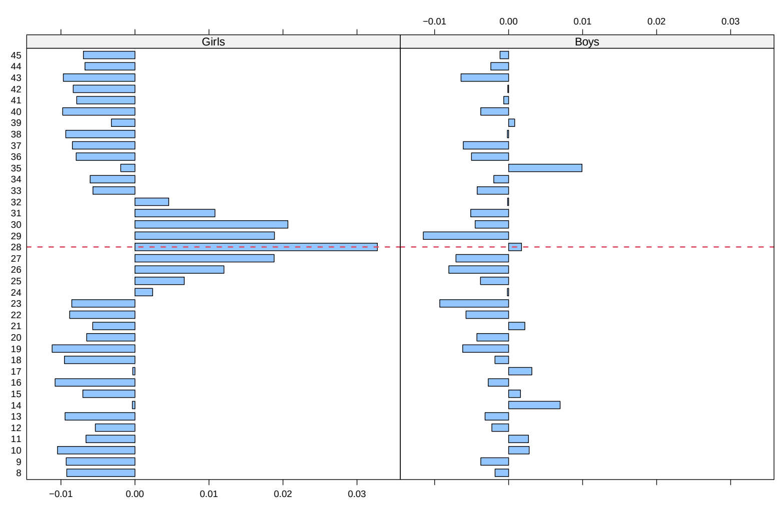
(a) 15%



(b) 20%



(c) 30%



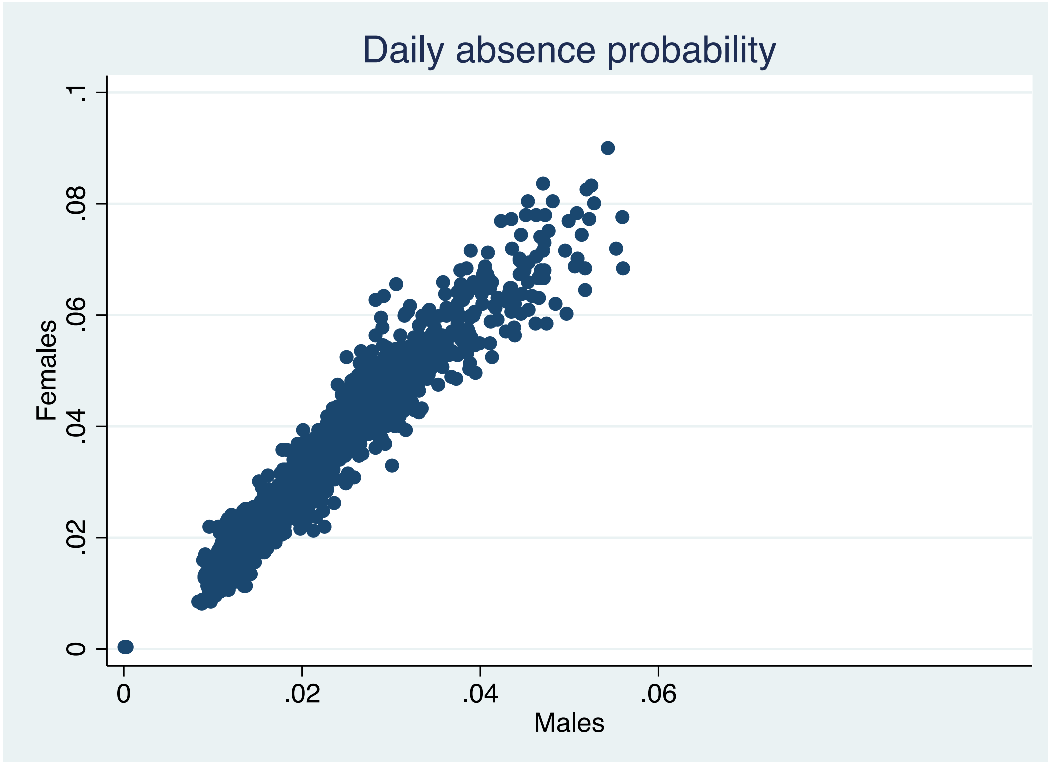


Figure 11: Daily absence correlation for Employee data

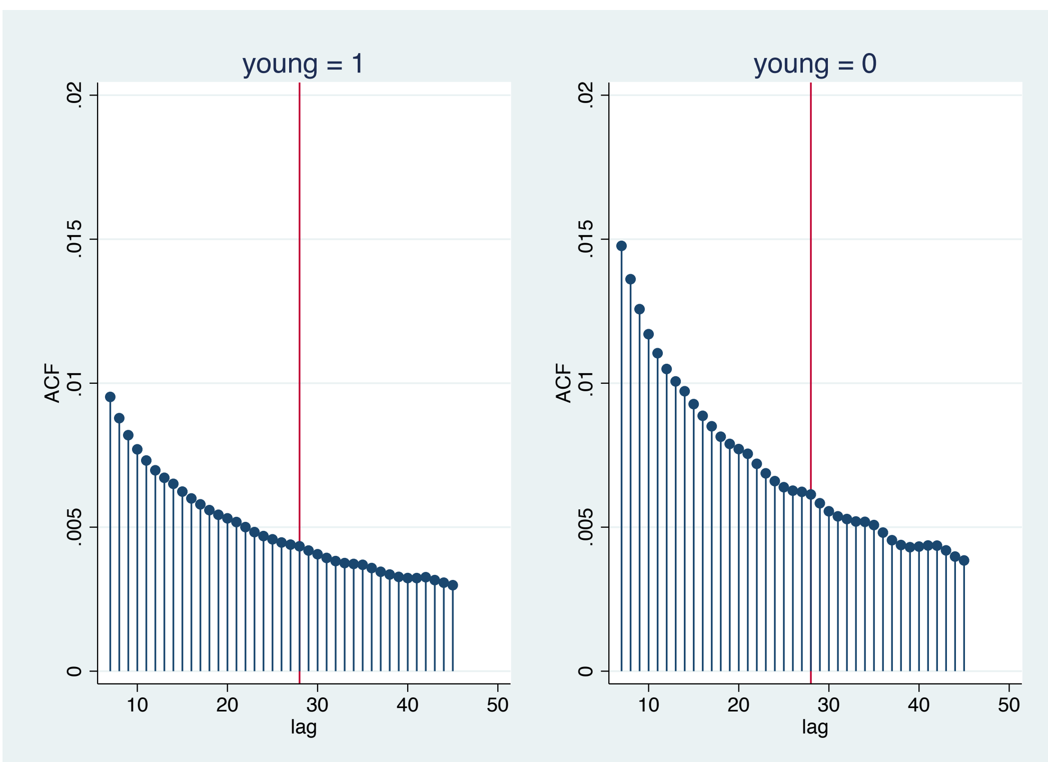


Figure 12: Aggregated autocorrelation function for Employee data

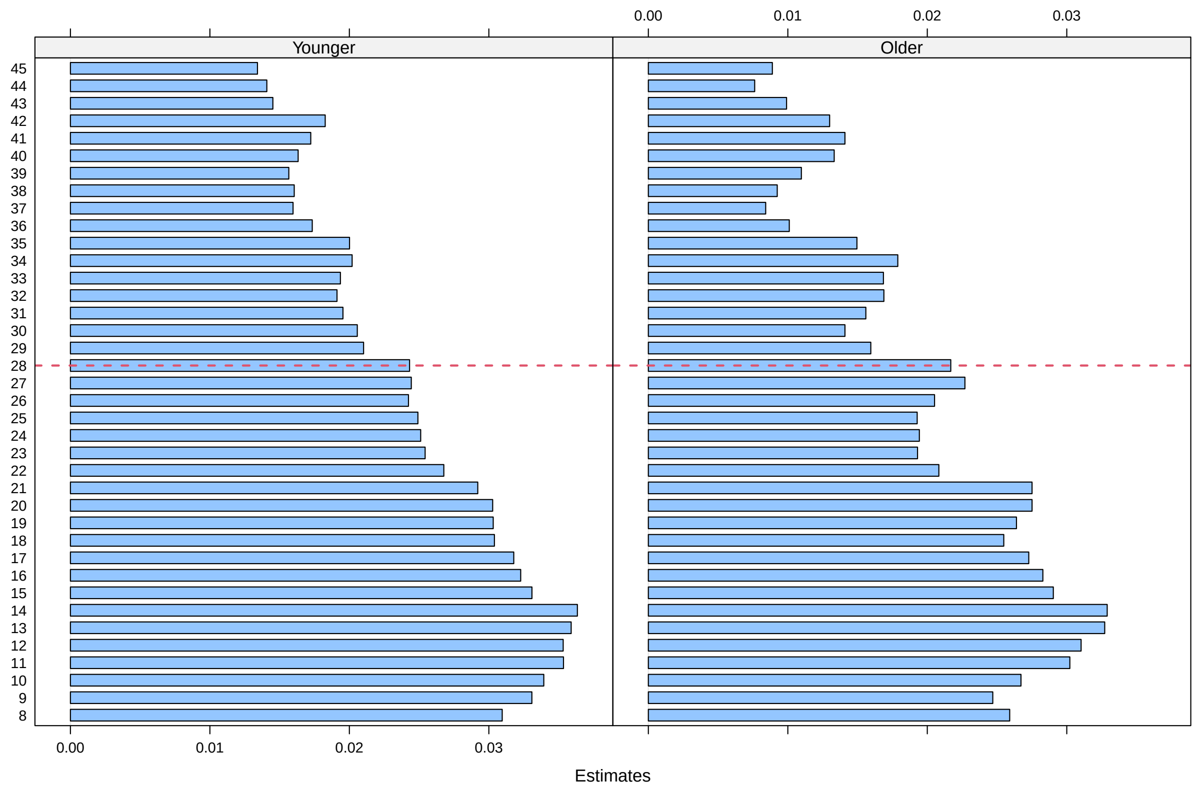


Figure 13: Regression coefficients for $lagX$ for Employee data

References

Agarwal, Garima. 2022. "Absenteeism, Menstruation and School-Level Programs - Findings from a Primary Survey in Delhi." *Unpublished Manuscript*.

———. 2025. "Sanitary Napkin Distribution and Girls' School Participation - Evidence from a Public Program in India." *Unpublished Manuscript*.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126 (4): 1709–53.

BBC News. 2017. "Kenya's Schoolgirls to Get Free Sanitary Pads from Government." https://www.bbc.com/news/world-africa-40365691.

Benshaul-Tolonen, Anja, Zulaika Garazi, Elizabeth Nyothach, Clifford Oduor, Linda Mason, David Obor, Kelly T Alexander, Kayla F Laserson, and Penelope A Phillips-Howard. 2019. "Pupil Absenteeism, Measurement, and Menstruation: Evidence from Western Kenya." *CDEP-CGEG Working Paper* 74.

Euro News. 2022. "Painful periods? Spain just passed Europe?s first paid 'menstrual leave' law," May. https://www.euronews.com/next/2023/02/16/spain-set-to-become-the-first-european-country-to-introduce-a-3-day-menstrual-leave-for-wo.

Grant, Monica, Cynthia Lloyd, and Barbara Mensch. 2013. "Menstruation and School Absenteeism: Evidence from Rural Malawi." *Comparative Education Review* 57 (2): 260–84.

Herrmann, Mariesa A, and Jonah E Rockoff. 2012. "Does Menstruation Explain Gender Gaps in Work Absenteeism?" *Journal of Human Resources* 47 (2): 493–508.

Hillard, Paula J Adams. 2002. "Menstruation in Young Girls: A Clinical Perspective." *Obstetrics & Gynecology* 99 (4): 655–62.

———. 2008. "Menstruation in Adolescents: What's Normal, What's Not." *Annals of the New York Academy of Sciences* 1135 (1): 29–35.

Ichino, Andrea, and Enrico Moretti. 2009. "Biological Gender Differences, Absenteeism, and the Earnings Gap." *American Economic Journal: Applied Economics* 1 (1): 183–218.

Kalia, Saumya. 2023. "Menstrual hygiene facilities in Indian schools." *The Hindu*. https://www.thehindu.com/sci-tech/health/explained-menstrual-hygiene-facilities-in-indian-schools/article66754551.ece.

Khadgawat, Rajesh, RK Marwaha, Neena Mehan, Vineet Surana, Aashima Dabas, V Sreenivas, M Ashraf Gaine, and Nandita Gupta. 2016. "Age of Onset of Puberty in Apparently Healthy School Girls from Northern India." *Indian Pediatrics* 53 (5): 383–87.

Krenz, Astrid, and Holger Strulik. 2019. "Menstruation Hygiene Management and Work Attendance in a Developing Country." *CEGE Discussion Papers, Number*.

Oster, Emily, and Rebecca Thornton. 2011. "Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 3 (1): 91–100.

Pathak, Praveen Kumar, Niharika Tripathi, and SV Subramanian. 2014. "Secular trends in menarcheal age in India - Evidence from the Indian Human Development Survey." *PloS One* 9 (11): e111027.

Scottish Government. 2018. "Students to Get Free Access to Sanitary Products." https://www.gov.scot/news/students-to-get-free-access-to-sanitary-products/.

UK Government. 2023. "How the Period Products Scheme is removing barriers in education." https://educationhub.blog.gov.uk/2023/09/28/how-the-period-products-scheme-is-removing-barriers-in-education/.

Van Biljon, Chloé, and Cobus Burger. 2019. "The Period Effect: The Effect of Menstruation on

Absenteeism of School Girls in Limpopo."
Vashisht, Aditi, Rambha Pathak, Rashmi Agarwalla, Bilkish N Patavegar, and Meely Panda. 2018. "School Absenteeism During Menstruation Amongst Adolescent Girls in Delhi, India." *Journal of Family & Community Medicine* 25 (3): 163.