Biological Gender Differences, Absenteeism and Earnings

Andrea Ichino
EUROPEAN UNIVERSITY INSTITUTE

Enrico Moretti
UC BERKELEY AND NBER

Table 1: Gender Differences in Days of Absence in a Year, by Country

	All workers		Unmarried and no chldren	
	Unconditional	Conditional	Unconditional	Conditional
	(1)	(2)	(3)	(4)
Continental Europe	6.67**	7.65**	2.12**	2.78**
	(0.52)	(0.60)	(0.80)	(0.88)
United Kingdom	11.17**	12.31**	2.28	2.20
	(2.10)	(2.53)	(2.45)	(2.68)
Usa	3.07**	3.09**	1.09**	2.01**
	(0.23)	(0.43)	(0.49)	(0.88)
Our Sample	4.66**	5.04**	2.76**	3.70**
	(0.32)	(0.33)	(0.53)	(0.54)

Introduction

Absenteeism is higher among female workers.

This is true even if we consider only illness-related episodes (excluding maternity-related absences).

In this paper we find that:

- part of this gender difference in absenteeism may be attributed to a *biological* difference between males and females;
- this biological difference has non-trivial earning consequences for females.

Result 1: Cyclical Absences

We use information on the exact date and duration of absence episodes in the personnel dataset of a large Italian bank.

We find that:

- the probability of an absence due to illness increases for females, relative to males, between the 26th and 30th day after the previous illness;
- the spike in the difference is at 28 days;
- the difference in the probability is more evident for young workers.

These results suggest that menstrual cycles induce an increase in the hazard of an absence episode.

Furthermore. the increase in the hazard of an absence due to the 28-days cycle:

- is higher among workers in line for merit promotions;

- does not change according to average local absenteeism;

These findings suggest that real pain may be an important cause of the cycle.

28-days cyclical absences explain a large fraction of the gender gap in total days lost for absences

Result 2: A Simple Model of Absenteeism and Earnings

Men are absent from work because of health and shirking shocks. Women face an additional source of absenteeism: menstrual cramps.

Menstrual cramps add noise to the determinants of absenteeism.

Thus signal extraction based on absenteeism is more informative about shirking for males than for females.

Consistent with the predictions of the model we find that

- The relationship between measures of worker quality and absenteeism is more negative for males
- The relationship between earnings and absenteeism is more negative for males
- This difference in slope declines with tenure

Result 3: What is the Effect on the Gender Earnings Gap?

In terms of earnings, we find that:

- in our sample, females earn 13.5% less than males;
- higher absenteeism induced by 28-days cycles may explain
  - 1.7 percentage points or
  - 13 percent

  of this earnings differential.

Data

- We only include full time employees who are on the payroll continuously for 3 years. We exclude workers in maternity leave.

- Our analysis is restricted to absence episodes due to illness;

- Workers can have an (almost) unlimited number of illness absences;

- Workers have to send a medical certificate if absence longer than 3 days. The certificate is easy to obtain;

Estimating 28-Days Periodicity

We test for the existence of a 28-days cycle in absenteeism:

- Analysis of the distance between all pairs of absence episodes.
- Duration analysis:
  - non-parametric hazard estimates;
  - parametric hazard estimates.

The two approaches generate consistent results.

Figure 1: Gender Differences in the Distribution of the Distance Between Consecutive Absence Spells.

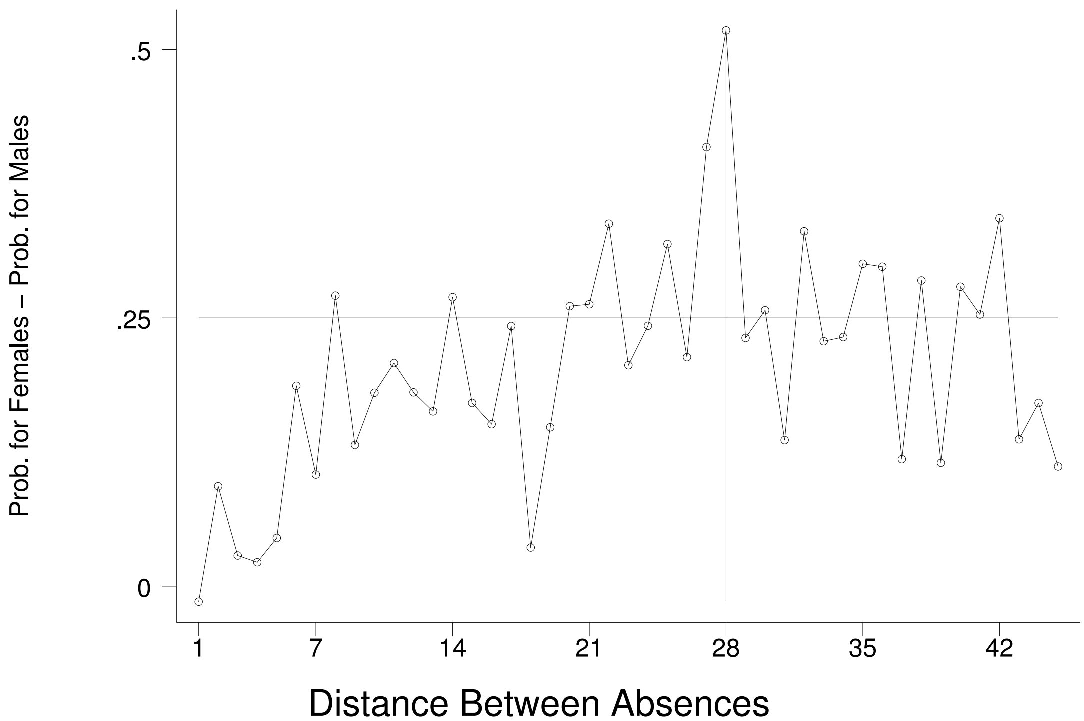
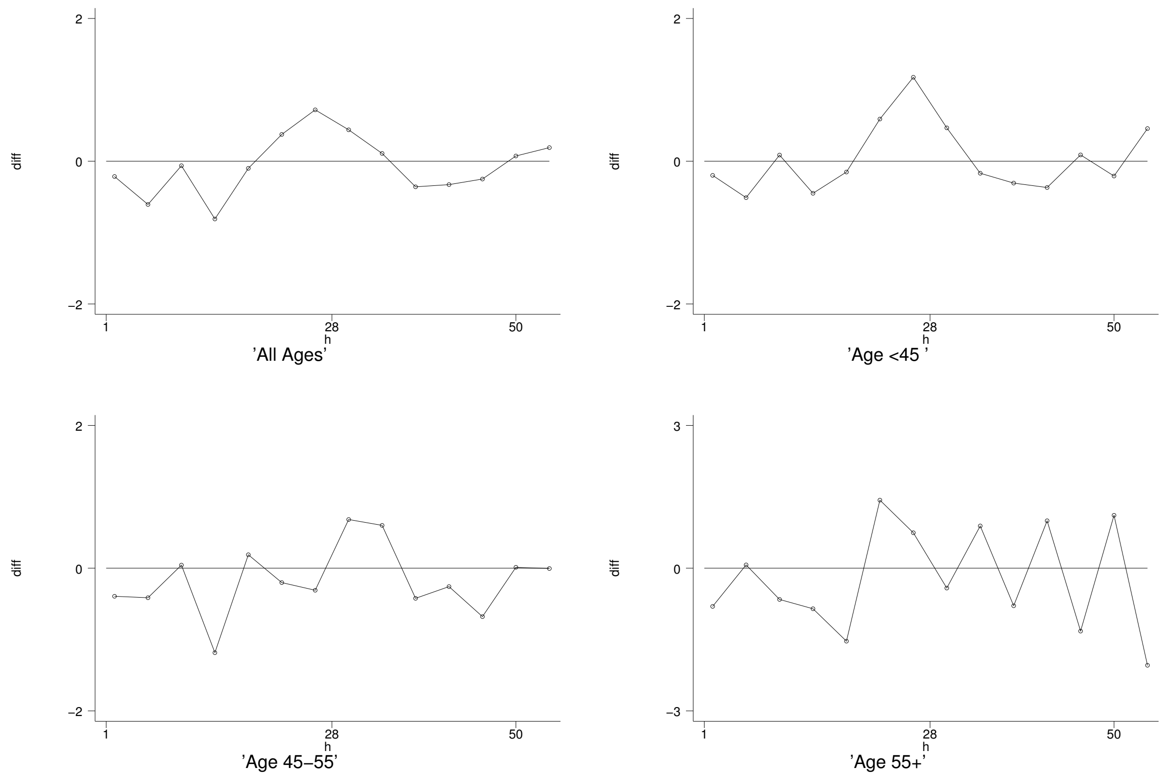


Figure 2: Gender Differences in the Distribution of the Distance Between Absence Pairs, Using All Possible Pairs.



Notes: Kaplan-Meier estimates of the hazard of a second absence episode.

Figure 3: Hazard Rates, by Gender

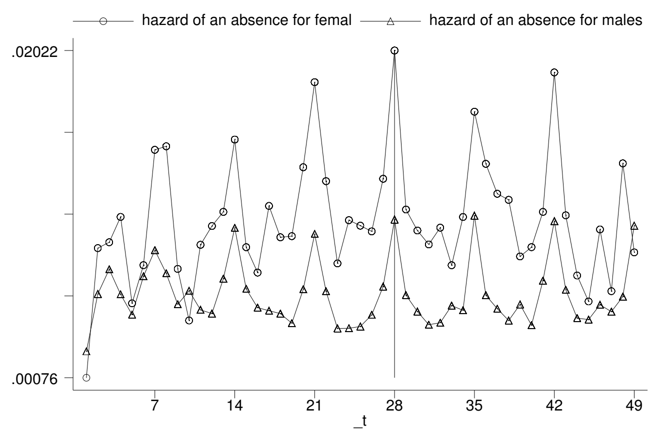


Fig. 1a: female and male hazard rates – all ages

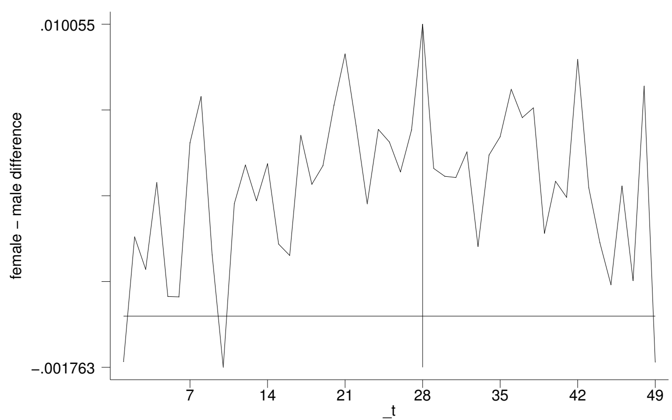


Fig. 1b: difference between female and male hazard rates – all ages

Notes: Kaplan-Meier estimates of the hazard of a second absence episode.

Figure 4: Difference in Hazard Rates, by Age

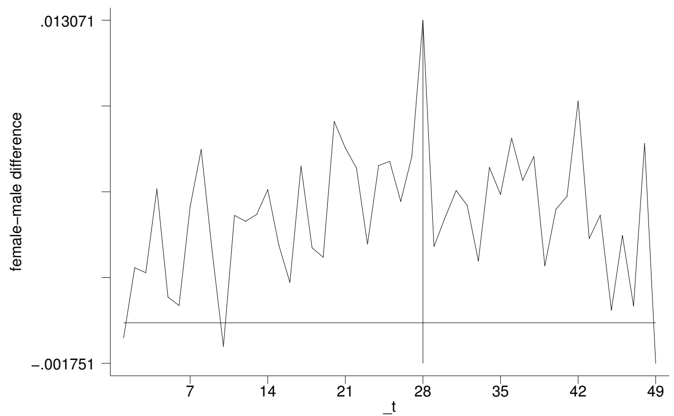


Fig. 2b: difference between female and male hazard rates − under 45

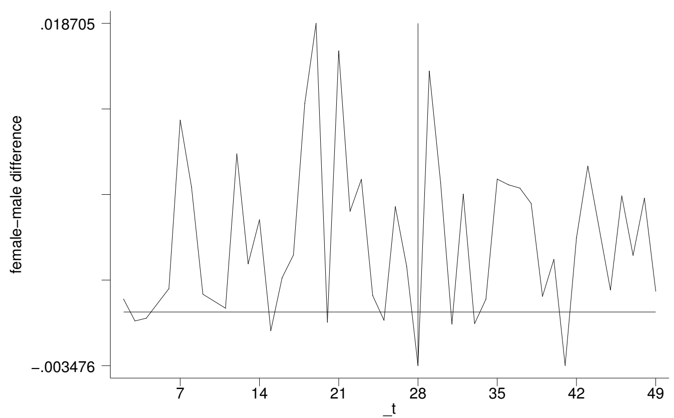


Fig. 3b: difference between female and male hazard rates − above 45

Notes: Kaplan-Meier estimates of the hazard of a second absence episode.

The 7-days periodicity

For both genders, the most likely weekday in which an absence begins is Monday;

There is a 7-days periodicity independently of the Monday effect. It is probably due to family and other non-work related commitments.

Since 28 is a multiple of 7, the 7-days periodicity creates a confounding pattern with respect to the pattern potentially induced by menstrual cycles.

Parametric estimates allow us to control appropriately for this confounding effect in order to test for the existence of a 28-days periodicity.

Parametric hazard estimates

$$h(t, X_{it}, \Psi) = \lambda(t)e^{\alpha+\beta F_i+\gamma M_{it}F_i+\delta S_{it}F_i} \qquad (1)$$

where:

- $X_{it} = (F_i, M_{it}, S_{it})$;
- $\Psi = (\alpha, \beta, \gamma, \delta)$;
- $\lambda(t)$ is the baseline hazard, which takes care of the 7-day periodicity;
- $F_i = 1$ indicates that i is a female;
- $M_{it} = 1$ if $t \in [26, 30]$:
  - i.e. whatever the baseline hazard, females can have a higher hazard between 26 and 30 days;
- $S_{it} = 1$ if $t = 7$ or multiples of 7:
  - i.e. females can have a different 7-day periodicity.

Table 2: Hazard of an absence for females relative to males and the risk of a menstrual cycle.

	Days at risk of menstrual cycle: $t \in [26, 30]$		Days at risk of menstrual cycle: $t = 28$		Days at risk of menstrual cycle $t = 28$ or 56	
	e^{β}	e^{γ}	e^{β}	e^{γ}	e^{β}	e^{γ}
All workers	1.58	1.23	1.59	1.15	1.59	1.38
N = 14857	(20.29)	(2.00)	(20.84)	(0.75)	(20.84)	(2.07)
Under 45	1.56	1.35	1.57	1.49	1.57	1.58
N = 10793	(17.89)	(2.57)	(18.47)	(1.89)	(18.47)	(2.65)
Above 45	1.58	1.07	1.58	0.35	1.58	0.66
N = 4064	(7.67)	(0.29)	(7.82)	(-1.67)	(7.82)	(-0.95)
Under 45, not monday	1.53	1.38	1.55	1.82	1.55	1.96
N = 7033	(14.15)	(2.32)	(14.75)	(2.07)	(14.75)	(2.78)
Above 45, not monday	1.57	1.03	1.58	0.22	1.58	0.55
N = 2668	(6.19)	(0.12)	(6.30)	(-1.44)	(6.30)	(-0.91)
Under 45, controlled	1.56	1.35	1.58	1.49	1.58	1.58
N = 10793	(17.08)	(2.53)	(17.61)	(1.88)	(17.61)	(2.65)
Above 45, controlled	1.42	1.06	1.43	0.35	1.43	0.66
N = 4064	(5.77)	(0.26)	(5.89)	(-1.67)	(5.89)	(-0.95)

Interpreting the 28-Day Periodicity

Two possible reasons for association between menstrual cycles and the hazard of an absence episode:

- menstrual pain may be so strong to induce an unavoidable absence;
- because of a social norm, an absence may be more justifiable for a female if it is associated with menstruations.

We restrict the sample to workers who are in line for a promotion (less likely to shirk). We find that:

- the hazard of an absence episode for females more than doubles, relative to males, during the days at risk of a menstrual cycle;
- this effect is larger than the one observed for all workers.

In branches with many females or located in the South we find that:

- both these environments are associated with higher absenteeism on average, but
- the hazard of an absence during the days of a potential menstrual cycle is not more evident for females working in these branches.

This evidence supports the hypothesis that the 28-days cycle is mainly caused by the real pain of menstruations.

Table 3: Hazard of an absence for females relative to males and the risk of a menstrual cycle at the top of the firm's hierarchy

Days at risk of menstrual cycle: $t \in [26, 30]$	e^{β}	e^{γ}
All workers	1.44	1.35
N = 5653	(7.27)	(1.15)
Under 45	1.42	2.06
N = 3302	(5.93)	(2.46)
Above 45	1.45	0.39
N = 2351	(3.84)	(-1.28)
Under 45, not monday	1.38	1.85
N = 2169	(4.41)	(1.79)
Above 45, not monday	1.39	0.29
N = 1543	(2.65)	(-1.18)
Under 45, controlled	1.46	2.03
N = 3302	(6.16)	(2.42)
Above 45, controlled	1.34	0.38
N = 2351	(2.90)	(-1.30)

Table 4: The role of different work environments

	Female environment		Southern environment	
	Below 45	Above 45	Below 45	Above 45
Hazard ratio (relative to males) for:				
Females (e^{β})	1.53	1.42	1.57	1.43
	(16.18)	(5.73)	(17.30)	(5.86)
Females in days at risk (e^{γ})	1.56	1.08	1.24	0.91
	(2.07)	(0.15	(1.59)	(-0.30)
Females in days at risk interacted with fraction of females in branch (e^{ϕ})	0.52 (-0.81)	0.92 (-0.04)		
Fraction of females in branch (e^{ψ})	1.52	0.99		
	(4.21)	(-0.05)		
Females in days at risk interacted with working in the south (e^{ϕ})			1.29 (1.36	1.56 (0.32)
Working in the south (e^{ψ})			1.27	1.35
			(10.59)	(7.91)
Controls	yes	yes	yes	yes
Number of workers	10793	4064	10793	4064

A Simple Model of the Relationship Between Earnings and Absenteeism

- If employers cannot directly observe individual productivity, they might use observables, including absenteeism, to set wages.
- Men are absent because of health shocks and shirking. Women might also suffer from menstrual cramps.
- Signal extraction based on absenteeism is more informative about productivity for men than for women (if menstrual cramps are not a signal of shirking.)
- Predictions:
  - i. The relationship between earnings and absenteeism should be more negative for men than for women.
  - ii. The relationship between worker quality and absenteeism should also be more negative for men than for women.
  - iii. These gender differences in slope should decline with worker tenure.

We adapt the career concerns model (Holmstrom, 1999).

Output depends on absenteeism:

$$Y_{it} = Q - X_{it} \tag{2}$$

$$X_{it} = P_{it} - e_{it} \tag{3}$$

P_{it} captures health shocks and "shirking" shocks.

e_{it} is effort

$$P_{it} = A_i + \epsilon_{it} \tag{4}$$

A_i is an individual specific time-invariant component (worker's quality). ϵ_{it} captures short run deviations.

We assume that

$$A_i \sim N(\mu_i, \frac{1}{h}) \tag{5}$$

$$\epsilon_{it} \sim N(0, \frac{1}{k_i}). \tag{6}$$

In other words, P_{it} is distributed normally with a mean and a variance that differ across genders.

We assume that the firm observes past absenteeism (or, equivalently, past output) but does not observe A_i, nor contemporaneous e_{it} and ϵ_{it}.

In period 2, the firm offers a wage equal to expected output given history.

$$W_{i2} = E(Y_{i2}|X_{i1}) \tag{7}$$

In a two period setting, $e_{i2} = 0$

$$E(Y_{i2}|X_{i1}) = Q - E(A_i|X_{i1}) \tag{8}$$

The firm predicts worker's quality using the normal learning model:

$$E(A_i|X_{i1}) = \frac{h}{h+k_i}\mu_i + \frac{k_i}{h+k_i}(X_{i1} + e^*_{i1}) \tag{9}$$

The worker chooses effort to maximize

$$\text{Max } [w_{i1} - C(e_{i1})] + E[W_{i2} - C(e_{i2})] \tag{10}$$

If $C(e_{it}) = \frac{\theta e_{it}^2}{2}$, optimal effort is

$$e_{i1}^* = \frac{k_i}{\theta(h + k_i)} \tag{11}$$

The equilibrium wage in period 2 is:

$$W_{i2} = \alpha - \beta X_{i1} \tag{12}$$

with

$$\alpha = Q - \frac{h}{h + k_i}\mu_i + \frac{k_i^2}{\theta(h + k_i)^2} \tag{13}$$

and

$$\beta = \frac{k_i}{h + k_i} \tag{14}$$

Gender differences

For women P is the result of two types of factors:

- shirking and health factors that have nothing to do with menstrual cramps;
- factors related to menstrual cramps.

$A_i = A_{1i} + \mu_2$

$\epsilon_{it} = \epsilon_{1it} + \epsilon_{2it}$

where ϵ_{2it} is orthogonal with respect to ϵ_{1it}

Men only have the first set of factors.

$A_i = A_{1i}$

$\epsilon_{it} = \epsilon_{1it}$

Implications

Average potential absenteeism is higher for females:

$$E(P_{it}|female) = \mu_1 + \mu_2 > \mu_1 = E(P_{it}|male) \quad (15)$$

The variance of potential absenteeism is higher for females:

$$Var(P_{it}|female) = \frac{1}{h} + \frac{1}{k_1} + \frac{1}{k_2} > \frac{1}{h} + \frac{1}{k_1} = Var(P_{it}|male). \quad (16)$$

⇒ Because of this difference in variance, absenteeism is a more noisy signal of worker quality for women than for men.

⇒ The slope in the earning equation is more negative for males:

$$\beta_m > \beta_f \quad (17)$$

The intercept in the earning equation is higher for males:

$$\alpha_m > \alpha_f \quad (18)$$

Extension to more than two periods

Over time, the firm learns more about workers. The fact that observed absenteeism is a more noisy measure for females becomes less relevant.

Iterating the Normal Learning Equation beyond two periods:

$$E(A_i|X_{i1},...,X_{it}) = \frac{h}{h+tk_i}\mu_i + \frac{k_i}{h+tk_i}\sum_{s=1}^{t}(X_{is}+e_{is}^{*}) \tag{19}$$

and the wage offer in period $t+1$ is

$$W_{it+1} = Q - \frac{h}{h+tk_i}\mu_i + \frac{k_i}{h+tk_i}\sum_{s=1}^{t}e_{is}^{*} - \frac{k_i}{h+tk_i}\sum_{s=1}^{t}X_{is} \tag{20}$$

The precision of the prior on A improves for both genders, until A becomes fully known.

In a regression of earnings on absenteeism, the coefficient decreases with tenure (in absolute value). Gender differences in slope decline.

Empirical Predictions of the Model

- In a regression of earnings on absenteeism the coefficient is negative and smaller in absolute value for females than for males.
- In a regression of measures of worker quality on absenteeism the coefficient is positive and smaller in absolute value for females than for males.
- In a regression of wages on absenteeism the coefficient decreases in absolute value with tenure, becoming more similar across genders.

Quantifying the Number of Cyclical Absences

For each worker, we want to estimate what fraction of absences has a cycle of 28 days.

We consider the distance between all pairs of short absences (3 days or less).

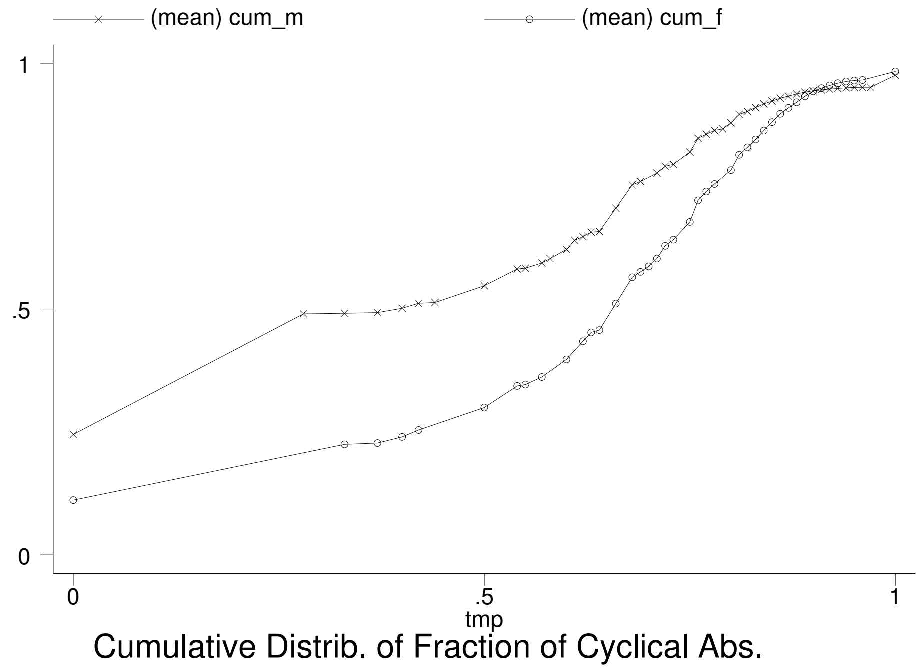
We call two absences cyclical if their distance is between 26 and 30 days or multiples.

We then normalize this number by the number of short absences so that we obtain an index ranging between 0 and 1.

Note: even for men this indicator may be larger than 0 because of

- false positive;
- 7-days cycle.

Figure 5: The Fraction of Cyclical Absences of Men stochastically Dominates the Fraction of women



Notes: the top line is the cumulative distribution of the fraction of cyclical absences for men. The bottom line is the cumulative distribution of the fraction of cyclical absences for women. Sample includes workers 45 or younger

Table 5: Absenteeism, by Type and Gender

	Men	Women		Difference	
			Unconditional	Conditional	Conditional
	(1)	(2)	(3)	(4)	(5)
Days of Absence					
Total Number of Days in a Year	8.2	12.9	4.6	5.2	5.4
			(.3)	(.3)	(.3)
Number of Cyclical Days in a Year	1.3	2.9	1.4	1.5	1.5
			(.06)	(.06)	(.06)
Episodes of Absence					
Total Number of Episodes in a Year	2.1	3.6	1.5	1.6	1.6
			(.5)	(.5)	(.6)
Number of Cyclical Episodes in a Year	.9	2.0	1.1	1.1	1.1
			(.04)	(.04)	(.04)
Control for Age			N	Y	Y
Control for Education			N	N	Y

Table 6: Number of cyclical episodes in a year, by Gender

Number of cyclical episodes	% Frequency for males	% Frequency for females
0	55	29
1	20	22
2	10	15
3	5	9
4	3	6
5	2	5
6	1	4
7	1	2
8	0	2
9	0	1
10+	1	4

Notes: sample includes workers 45 or younger.

Medical Literature

- A recent study by Chawla et al. (Medical Care, 2002) aims at measuring the "economic burden" of premenstrual dysphoric disorder (PMDD).
- It uses a representative sample of 1194 women aged 21 to 45 in California.
- Its estimates suggest that on average PMDD causes 2.4 days of absence from work annually.
- This number is slightly larger than our estimates.

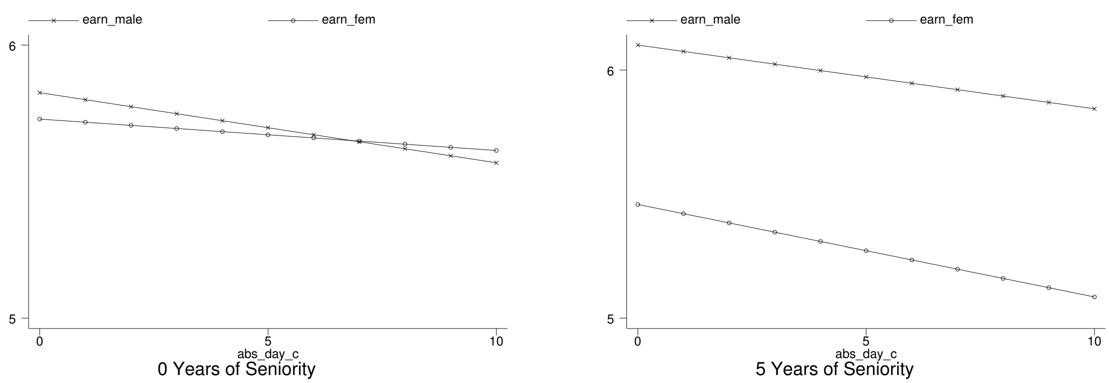
Table 8: Earnings and Career Equations - Linear Models

	(1)	(2)	(3)
Model 1: Earnings			
Female	-.204	-.135	-.144
	(.006)	(.006)	(.008)
Cyclical Absences			-.025
			(.001)
Female × cyclical absences			.010
			(.002)
Model 2: Promoted to Manager			
Female	-.183	-.111	-.138
	(.009)	(.009)	(.012)
Cyclical Absences			-.029
			(.002)
Female × cyclical absences			.017
			(.003)
Model 3: 6 Occupation Levels			
Female	-.273	-.136	-.204
	(.017)	(.017)	(.022)
Cyclical Absences			-.061
			(.003)
Female × cyclical absences			.041
			(.005)
Model 4: 13 Occupation Levels			
Female	-.754	-.216	-.368
	(.050)	(.047)	(.060)
Cyclical Absences			-.184
			(.000)
Female × cyclical absences			.108
			(.015)
Model 5: Merit Promotion			
Female	-.050	-.027	-.030
	(.008)	(.008)	(.011)
Cyclical Absences			-.013
			(.001)
Female × cyclical absences			.005
			(.002)
Controls for non-cyclical absences	N	Y	Y
Controls for age	N	Y	Y

Table 9: The Relationship Between Earnings and Cyclical Absences, by Gender and Firm Seniority

	Dependent Variable is Earnings (1)
Female	-.035
	(.016)
Cyclical Absences	-.025
	(.001)
Seniority	.0160
	(.001)
Female × Cyclical Absences	.018
	(.004)
Female × Seniority	-.010
	(.001)
Female × Cyclical Absences × Seniority	-.0005
	(.0002)

Figure 6: The Relationship Between Predicted Earnings and Cyclical Absences, by Gender and Firm Seniority



Note: The lines show the predicted log earnings by gender based on estimates of the model shown in Table 9.

Table 7: Gender Differences in Observable Indicators of Workers Quality, by Amount of Cyclical Absences

	Schooling	Sanctions	Days of Vacation Taken	Days of Strike
	(1)	(2)	(3)	(4)
Average Dep. Variable	13.1	.09	19.5	.97
Model 1: Linear				
Days of Cyclical Absence	-.104	.028	.018	.002
	(.014)	(.001)	(.015)	(.003)
Days of Cyclical Abs. × Female	.094	-.018	-.029	-.000
	(.023)	(.002)	(.025)	(.005)
Model 2: Three Groups				
Medium number of cyclical absences	-.182	.012	-.034	.018
	(.081)	(.009)	(.088)	(.020)
High number of cyclical absences	-.547	.074	.067	.031
	(.081)	(.009)	(.088)	(.019)
Small number of cyclical abs.× female	.165	-.019	-.066	.078
	(.142)	(.017)	(.152)	(.033)
Medium number of cyclical abs.× female	.488	-.018	.054	.043
	(.138)	(.008)	(.151)	(.032)
High number of cyclical abs. × female	.648	-.063	-.047	.092
	(.105)	(.012)	(.115)	(.024)
Test coeff on High < 0 (p-value)	.00			
Test coeff on High × Fem. > coeff on Low × Fem. (p-value)	.00			
Test coeff on High > 0 (p-value)		.00	.35	.05
Test coeff on High × Fem. < coeff on Low × Fem. (p-value)		.02	.92	.74
Controls	Y	Y	Y	Y

How much of the gender gap in earnings is explained by menstrual cycle?

Two possible answers:

(1) Back of the Envelope:

[Days of work lost due to cycle $\times$ Women average daily earnings] / Earnings gender gap

$$[1.5 \times (25020/230)]/4014 = 4\%$$

This estimate is likely to understate the true effect of cycle on earnings gap, because it does not reflect:

i. fixed costs

ii. lost productivity when on the job

iii. signaling value of avoiding absences

(2) Counterfactual Wage Gap:

We divide workers in three groups according to their number of cyclical absence episodes.

Men earnings

$$Y_m = \pi_{1m} Y_{1m} + \pi_{2m} Y_{2m} + \pi_{3m} Y_{3m} \tag{21}$$

Women earnings

$$Y_f = \pi_{1f} Y_{1f} + \pi_{2f} Y_{2f} + \pi_{3f} Y_{3f} \tag{22}$$

Observed earnings gap

$$Y_m - Y_f = (\pi_{1m} Y_{1m} - \pi_{1f} Y_{1f}) + (\pi_{2m} Y_{2m} - \pi_{2f} Y_{2f}) + (\pi_{3m} Y_{3m} - \pi_{3f} Y_{3f})$$

The counterfactual earnings gap assigns the distribution of males to females:

$$\tilde{Y}_m - \tilde{Y}_f = \pi_{1m}(Y_{1m} - Y_{1f}) + \pi_{2m}(Y_{2m} - Y_{2f}) + \pi_{3m}(Y_{3m} - Y_{3f}) \tag{23}$$

The difference between the observed and the counterfactual gap can be attributed to menstrual cycle, if the cycle is the only cause of the difference in the π's.

Estimation

$$logY_i = \beta_1 + \beta_2 C_{2i} + \beta_3 C_{3i} + \gamma_1 C_{1i} F_i + \gamma_2 C_{2i} F_i + \gamma_3 C_{3i} F_i + \mu X_i + e_i$$

F_i is an indicator for females;

C_{Ji} is an indicator for the j quantile of the cyclical absences distribution;

X_i controls for non-cyclical absences and age.

The parameters of interests are γ_j, the earnings gap *within each quantile*:

Key identifying assumption:

$E(e_i|F, j = 3) - E(e_i|M, j = 3) >= E(e_i|F, j = 1) - E(e_i|M, j = 1)$

the female-male difference in unobservables can not be negatively correlated with cyclical absences.

Note:

1) The assumption is about gender differences. We are *not* assuming that the unobservables can not be negatively correlated with cyclical absences.

2) The model predicts that the gender difference in workers quality should *increase* with day of cyclical absence.

Table 10: Earnings and Career Equations: Workers are Divided in Three Groups Based on the Number of Cyclical Absences

	Earnings	Manager	6 Occupation Levels	13 Occupations Levels	Merit Promotion
	(1)	(2)	(3)	(4)	
Medium number of cyclical absences	-.042	-.037	-.092	-.251	-.015
	(.007)	(.010)	(.019)	(.053)	(.010)
High number of cyclical absences	-.118	-.135	-.279	-.821	-.063
	(.007)	(.010)	(.019)	(.053)	(.009)
Small number of cyclical absences × female	-.131	-.107	-.173	-.302	-.033
	(.012)	(.019)	(.034)	(.092)	(.017)
Medium number of cyclical absences × female	-.118	-.116	-.138	-.222	-.019
	(.012)	(.018)	(.033)	(.091)	(.017)
High number of cyclical absences × female	-.099	-.059	-.012	.142	-.003
	(.009)	(.014)	(.025)	(.070)	(.013)
Test coeff on High < 0 (p-value)	.00	.00	.00	.00	.00
Test coeff on High × Fem. $>$ coeff on Low × Fem. (p-value)	.02	.02	.00	.00	.08
Observed Gender Gap (conditional)	-.135	-.111	-.136	-.216	-.027
Counterfactual Gender Gap (conditional)	-.119	-.096	-.119	-.161	-.022
Percent of the Observed Gap "Explained" by Cycle	11.9%	13.5%	12.5%	25.4%	18.5%

Conclusions

- As predicted by a simple model of asymmetric information, a cyclical absence "costs" more for males than females.
- However, females have more cyclical absences than males because of menstrual cycles.
- The (conditional) gender gap in earnings between males and females is 13.5%
- The gender gap if males and females had the male distribution of cyclicality would be 11.8%
- We conclude that if females did not have menstruations, the gender gap in earnings would be 13 percent lower.

Appendix Table A1: Descriptive statistics

	All		Absent at least once		Pairwise differences	
	Females	Males	Females	Males	Females	Males
Age	35.7	40.5	35.6	40.3	35.9	40.5
	(7.8)	(7.8)	(7.9)	(7.8)	(7.9)	(7.9)
Years of schooling	13.3	13.0	13.3	13.0	13.3	13.0
	(2.7)	(3.3)	(2.7)	(3.3)	(2.7)	(3.3)
Tenure	13.1	16.4	13.0	16.2	13.4	16.4
	(7.7)	(8.0)	(7.7)	(7.9)	(7.8)	(7.9)
Av. monthly wage (Euros)	1884	2590	1872	2537	1880	2551
	(560)	(1185)	(539)	(1133)	(541)	(1139)
Percent born in the south	28.4	35.5	28.7	36.2	28.0	35.6
Percent working in the south	25.5	28.3	25.7	28.9	25.0	28.3
Percent married	63.0	78.6	63.3	78.7	62.3	78.6
Percent with children	46.5	61.3	47.7	66.9	46.8	66.7
Percent manager	3.5	19.7	3.3	18.3	3.5	19.5
Percent high white collar	5.2	11.2	5.1	11.1	5.0	11.0
Percent low white collar	90.4	64.6	90.7	65.9	90.5	65.1
Percent high blue collar	0.7	3.4	0.7	3.6	0.7	3.5
Percent low blue collar	0.2	1.0	0.2	1.0	0.2	1.0
Number of observations	3040	13001	2965	11892	2822	12923
Percent of total	19.0	81.0	19.9	80.1	17.9	82.1

Note: Columns 1 and 2 of the table report statistics on workers continuously on payroll between January 1, 1993 and December 31, 1995 (workers on maternity leave during this period are excluded; see footnote **??**). In columns 3 and 4, the statistics refer to the workers who are absent at least once for illness related reasons during the same period. This is the sample of workers which is used later for the duration analysis. Finally, in column 5 and 6, statistics refer to the workers used in the analysis based on pairwise differences. The percent with children is based on the employees' declarations for tax deductions. The wage is computed as the annual nominal taxable income (in Euros, for year 1993) that the bank declares to have paid to the worker, divided by 12.

Appendix Table 2: Distribution by gender of the week-day in which an absence begins.

Week day in which absence begins	Females	Males	Total
Sunday	1	9	10
	(0.03)	(0.08)	(0.07)
Monday	972	4184	5156
	(32.78)	(35.18)	(34.70)
Tuesday	630	2366	2996
	(21.25)	(19.90)	(20.17)
Wednesday	527	1999	2526
	(17.77)	(16.81)	(17.00)
Thursday	482	1872	2354
	(16.26)	(15.74)	(15.84)
Friday	351	1438	1789
	(11.84)	(12.09)	(12.04)
Saturday	2	24	26
	(0.07)	(0.20)	(0.18)
Total	2965	11892	14857
	(100.00)	(100.00)	(100.00)

Note: The table reports statistics for the first absence episode of the workers continuously on payroll between January 1, 1993 and February 29, 1996 (workers on maternity leave during this period are excluded; see footnote **??**). Percent by columns in parentheses. Since some branches of the bank may occasionally open on weekends, absence episodes may also begin on Saturday or Sunday.

Appendix Table 3: Distribution of the fraction of absenteeism due to cyclical episodes, by Gender - Age Under 45

	Fraction in Group Who Are		Number of Cyclical Abs. Days		
	Males	Females	Mean	Min	Max
Bottom quantile (low cyclicality)	49	22	0	0	0
Middle quantile	23	26	1.1	1	1.6
Top quantile (high cyclicality)	28	52	4.5	2	33