Estimating Periodic Absences: Hazard Rates and Cyclical Data

Colin D. Sullivan[†]

Abstract

Efforts by labor economists to estimate the impact of cyclical health patterns on worker absences have yielded mixed results. In this research note I demonstrate that a tool commonly used in this domain, the Cox proportional hazards model, fails to detect the cyclical elements even in two ideal, simulated datasets. A new set of tools is needed to accurately estimate the impact of cyclical phenomena on labor outcomes such as absences and productivity.

I. Introduction

While it has been repeatedly observed that female workers have higher rates of absenteeism than males, the reasons for this disparity have not been adequately explained. Recent literature has attempted to attribute part of the difference to the effect of the menstruation cycle on women's willingness to attend work. To estimate the impact of a cyclical willingness to work economists have applied hazard models to data on absences of male and female workers. Ichino & Moretti (2009) investigated the attendance records of Italian bank employees, finding a significant 28-day effect for female workers. Herrmann & Rockoff (2011) disputed this finding after correcting programming errors and adapting the original model. They also applied the revised model to a new dataset of New York City teachers. They claimed that when the model is properly adjusted, no 28-day cycles are distinguishable in female absences. Both sets of authors apply the Cox proportional hazards (CPH) model on the time between consecutive absence spells to arrive at their findings. As I will show, however, this model is inadequate and inappropriate in this context, and will fail to identify the cyclical effect even in simple periodic datasets.

The distance between consecutive absence spells is unlikely to display the cyclical nature of the underlying absence probability function. I will show this below by presenting datasets derived from two types of cyclical absence functions that should display strong evidence of periodicity when appropriately analyzed. The CPH model finds no sign of this effect. This paper is organized as follows: in the next section, I present two datasets produced from ideal cyclical functions with a stochastic element; in Section III, I explain the analytical model used by both Ichino & Moretti and Rockoff & Hermann, and display the results of mis-applying this model in this context; and Section IV concludes.

[†]Harvard Kennedy School

II. Two Simple Periodic Datasets

To demonstrate the failure of the Cox proportional hazards model to detect periodic phenomena, we consider a simple model with male and female workers. All workers are expected to work every day, but are allowed absences. Every day, each worker's absence is determined by a Bernoulli trial with the probability of absence determined by a periodic function. I consider two different possible shapes for this function - one for Group 1 and another for Group 2 - below. The periodic functions are also designed to differ between men and women in each group - for men, probability of absence is given by a function with a seven day period only, while for women the probability depends on both a seven day period and an additional twenty-eight day period. The absence functions of both groups have been calibrated to be consistent with many of the stylized facts found in typical teacher attendance data - the total number of days observed is 180, and the average absence rate is around 5%, with women's absence rates a few percentage points higher than men's.

Group 1: Continuous (Sinusoidal) Absence Probability Function

For men in Group 1 the mean of the Bernoulli trial fluctuates periodically every seven days. The expected value of the trial for Group 1 on day x is given by

$$\frac{1}{50}\left(\frac{1}{2}\cos(2\pi x/7)+2\right).$$

The graph of the function is displayed below. Notice that it is periodic with a period of seven days, and oscillates between (min) and (max). For any full period, the average is 4%. This function describes the likelihood of absence for Group 1. That is, on day three, any member of Group 1 has a $\frac{1}{50}\left(\frac{1}{2}\cos(2\pi*3/7)+2\right)=4.62\%$ chance of absence.

For women, the mean of the Bernoulli trial still has a periodic element that lasts seven days, but it also contains a 28-day element. It is given by the following function:

$$\frac{1}{50}\left(\frac{1}{2}\cos(2\pi x/7)+\cos((2\pi x/28)+A_n)+2\right)$$

where A_n is a random number between 0 and 27 which offsets the phase of the 28-day element. That is, the mean of the Bernoulli trials for each woman looks like the function in Figure 2 shifted to the right by A_n.

Group 2: Rectangular (Discontinuous) Absence Probability Function

In Group 2, subjects' absence probability functions are rectangular rather than sinusoidal. The baseline probability for both men and women is 0.04 and jumps discontinuously at certain intervals. For both genders, the probability of absence is 2 percentage points higher every 7 days. For women, absence is raised a further 4 percentage points for 5 days at 28-day intervals, offset by a number of days chosen at random. The probability functions are displayed in Figures 3 and 4 below.

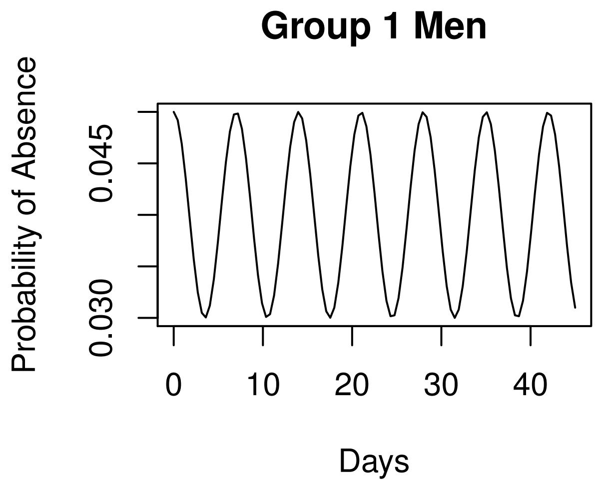

Figure 1: Likelihood of Absence for Group 1 Males

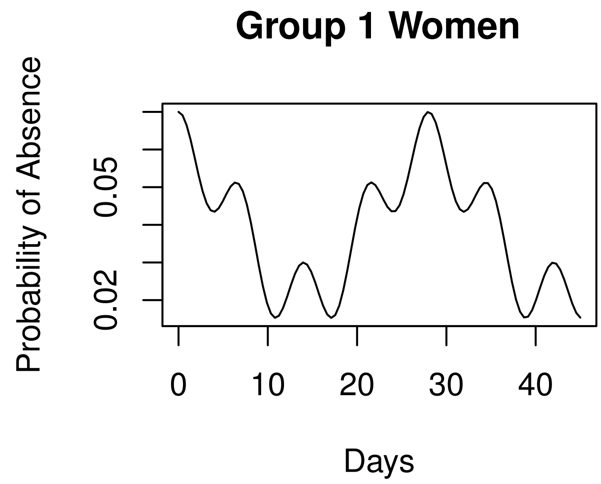

Figure 2: Likelihood of Absence for Group 1 Females

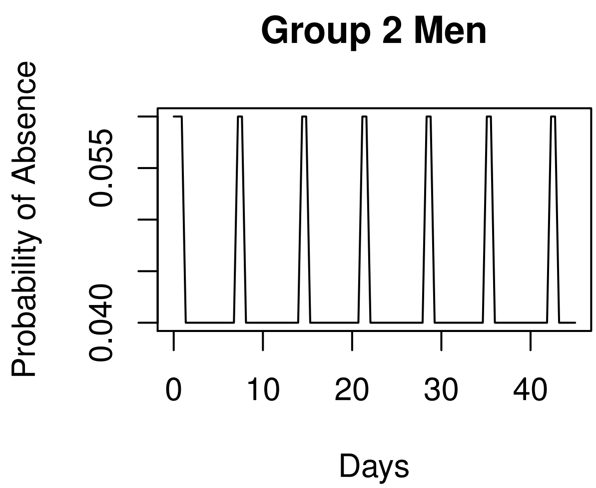

Figure 3: Likelihood of Absence for Group 2 Males

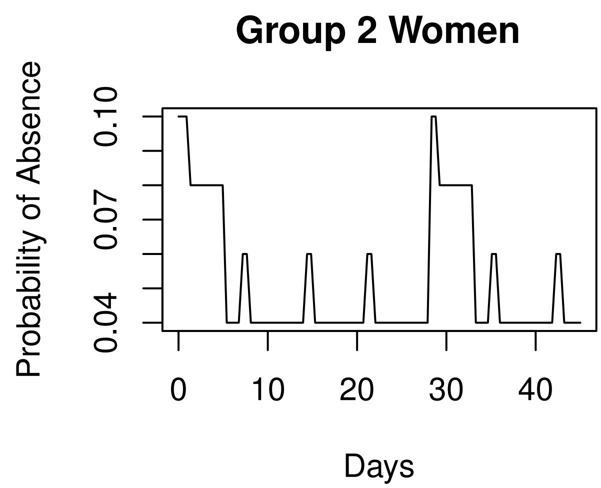

Figure 4: Likelihood of Absence for Group 2 Females

III. Model and Results

Using the absence functions described above, I simulated absence records for 1000 individuals over a period of 180 days and calculated the time between consecutive absence spells for each individual, as outlined in Hermann & Rockoff (HR) and Ichino & Moretti (IM). These records are the data for my analysis. Using the Cox proportional hazard model, I attempt to differentiate the impact of the 28-day cycle between the two groups using two models. Model 1 follows IM who found a significant effect of menstruation; Model 2 is a variant of the first based on HR.

Model 1

IM estimate the importance of the 28-day cycle by estimating the hazard for teacher i at time t controlling for whether the teacher is female (F_i), an interaction between female and 28 days ($F_i * M_{it}$), and an interaction between female and days that are multiples of seven ($F_i * S_{it}$). I have excluded IM's controls for the day of the week and worker characteristics:

$$h(t, X_{it}) = \lambda(t) exp(\alpha + \beta F_{it} + \gamma F_i M_{it} + \delta F_i S_{it})$$

where γ represents the "difference in the hazard rates of men and women 28 days after the start of the previous spell, after allowing for both a different baseline hazard (β) and seven-day periodicity (δ) for women."[1]

Model 2

HR revise the IM model to adjust for the effect of the occurrence of non-work days (weekends) at seven-day intervals. Since my data have no weekends by assumption, this adjustment should have no substantial effect on the results. The model allows the effect of distances that are a multiple of seven differ from one another by including a separate interaction variable between gender and each multiple of seven.

$$h(t, X_{it}) = \lambda(t) exp(\alpha + \beta F_{it} + \gamma F_i M_{it} + \delta_1 F_i(7 \text{ days}) + \delta_2 F_i(14 \text{ days}) + ...)$$

The regression results of Model 1 are displayed in the first and third columns of Table 1, for the datasets produced with sinusoidal and rectangular absence functions, respectively. Model 2 results are display in the second and fourth columns. The values of $e^{\text{coefficient}}$ are displayed for ease of interpretation: a value greater than one indicates a positive effect, while

[1] Herrman, Mariesa and Jonah Rockoff. 2011. "Does Menstruation Explain Gender Gaps in Work Absenteeism?" *NBER Working Paper No. 16523*: 3.

an value less than one should be interpreted as a negative effect. Below each coefficient value p-values are displayed in parentheses.

Estimates for a female 28-day effect derived from both the IM and HR models are between 0.5 and 0.9 - that is, both models from both datasets indicate a slightly *lower* likelihood of absence for females at 28-day intervals. These results are striking: the impact of the 28-day cycle is indistinguishable using the Cox proportional hazards model.

Table 1: Results of Cox Proportional Hazards Model

	Sinusoidal		*Rectangular*	
	Model 1	Model 2	Model 1	Model 2
female * 28-day	**0.821**	**0.751**	**0.884**	**0.571**
	(0.67)	**(0.53)**	**(0.78)**	**(0.17)**
female	0.997	0.967	1.070	1.04
	(0.96)	(0.61)	(0.31)	(0.59)
*female * multiple of 7*	0.887		0.629	
	(0.14)		(0.01)	
Separate variables for *female * multiples of 7*	No	Yes	No	Yes

The value of exp(coefficient) *is displayed with p-values in parentheses below. In bold is the estimate of interest,* female * 28-day. *This estimate is less than* 1 *in all cases, suggesting that the 28-day effect is indistinguishable using these models, or if anything that males are more likely to be absent at 28-day intervals.*

IV. Conclusion

The Cox proportional hazards model, which has been used in labor economics to estimate the impact of the menstruation cycle in worker absences, fails to capture cyclical effects even in idealized data sets. This paper contributes to the literature by identifying a key shortcoming of the CPH model in this context. Although this paper has focused on data that mirror worker absences, further research should address whether the model accurately estimates periodic effects for data with different characteristics, such as a higher expected absence rate. This would allow researchers to determine the effective limits of the CPH model and where a new model is needed. Datasets similar to those produced above could easily be created to answer these questions.

Furthermore, a new set of analytical tools must be developed to analyze periodic economic phenomena. Research to address this gap in the economic toolkit will be valuable for a range of behavioral questions that go beyond worker absences.

References

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

Herrmann, Mariesa and Jonah Rockoff. 2011. "Does Menstruation Explain Gender Gaps in Work Absenteeism?" *NBER Working Paper No. 16523.*