Subjective Performance Evaluation in a Multi-tasking Environment: a Firm-level Experiment in China

Thomas Gall
University of Southampton

Xiaocheng Hu
University of Exeter

Michael Vlassopoulos
University of Southampton

Cedefop-Eurofound-IZA
August 2020

Motivation

- How to incentivise line managers?
- Difficult, because
  - optimal reward system often needs to induce them both to do own work and to invest in leadership, organisational skill, etc.
  - performance on productive tasks relatively easy to measure and reward, but how to measure soft, managerial inputs?
- Many firms use subjective performance evaluation for hard to measure inputs and outputs (Bartel et al. 2017); surveys by LPP and GPCS find more than half of surveyed firms use subjective measures.
- To solve multitasking problem: combine subjective with objective performance measures of different dimensions?

This Paper

- Field experiment in factories using team production:
  - Foreman (line manager):
    i. Take part in producing output (well quantifiable).
    ii. Lead and organise their teams (hard-to-measure).
  - However, managerial input believed to be an underprovided public good (increasing group productivity).
- Treatment:
  - Subjective evaluation of foreman's management performance.
  - This is a relative performance evaluation.
  - Monetary rewards depend on (public) ranking.
- Results:
  - Team productivity increased by 5%, driven by group members.
  - Foremen's productivity did not change significantly, but they increased working hours.
  - Intervention is profitable for the firm.

Literature

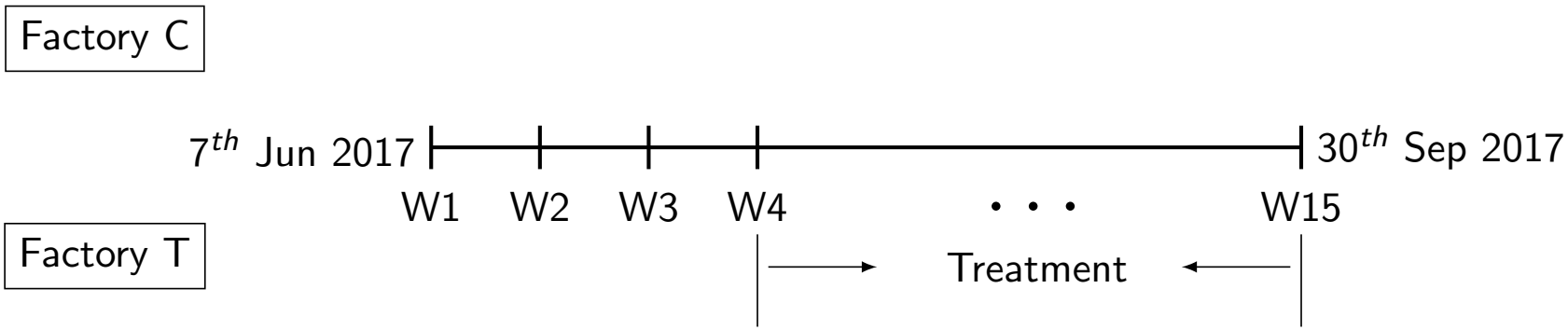
- Subjective evaluation:
  - e.g. Berger et al (2013), Frederiksena et al (2017), Grund and Przemeck (2012), Kampkötter and Sliwka (2018).
- Multitasking:
  - e.g. Al-Ubaydli et al. (2008), Coviello et al (2014, 2015), Dumont et al. (2008), Englmaier et al. (2014), Fryer and Holden (2012), Hong et al. (2018), Hossain and List (2012), Hossain and Li (2014), Johnson et al. (2012), Kishore et al. (2013), Kremer et al. (2010), Larkin (2014), Mullen et al. (2010), Paarsch and Shearer (2000), Shearer (2004).
- Subjective evaluation in a multitasking environment:
  - e.g. Bartel et al. (2017), Engellandt and Riphahn (2011); but: Bol (2011), Takahashi et al. (2014).
- This talk: a field experiment on productivity effects of adding subjective evaluation of soft, managerial inputs to output based pay in a multitasking environment.

The Setting

- 2 factories, assembly lines producing the same products.
- Each line consists of 1 forewoman and 3-10 workers.
- Forewomen and workers pack products, paid by piece rate.
- In addition, forewomen organise the production line, paid a flat rate (5% of monthly income).

Experimental Design

- 1 control factory (C, 27 employees) and 1 treatment factory (T, 43 employees).
- 15 experimental weeks.
- Since W2 the evaluation secretly took place in both factories.
- Treatments were introduced in factory T at the end of W3.



The Treatment - Subjective Evaluation

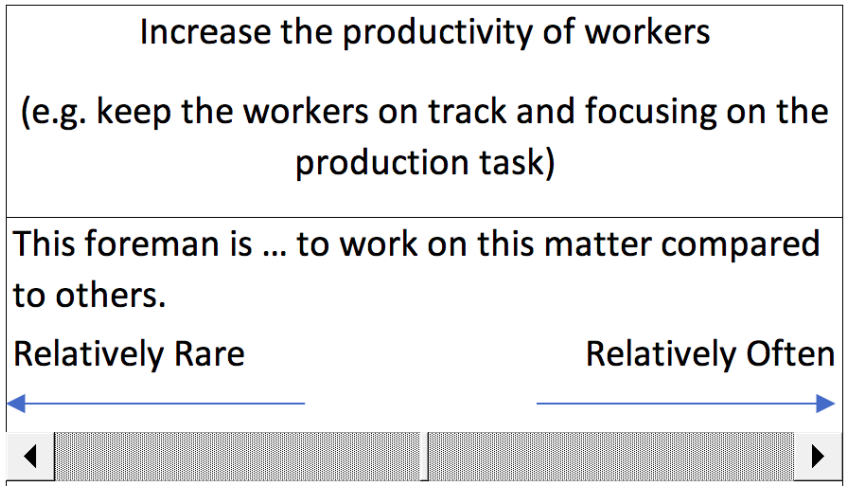
- Four organisational activities (inputs) are evaluated.
  - **Organisation**: maintain an efficient production process (e.g. make sure the raw materials are sufficient and unerring on the line for workers to work with).
  - **Productivity**: increase the productivity of workers (e.g. keep the workers on track and focusing on the production task).
  - **Quality**: reduce line defect rates (e.g. constantly remind workers to use the standardised operating procedure in order to reduce the number of faulty products).
  - **Relationship**: team building (e.g. provide support and communication to foster a friendly and positive work environment).

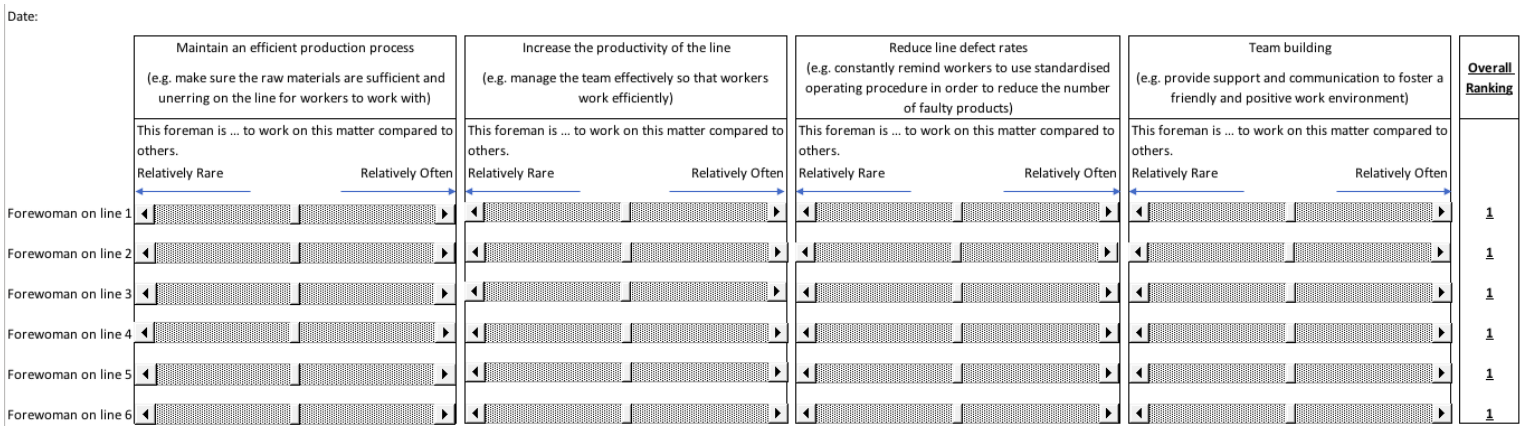
The Treatment - Subjective Evaluation

- Four organisational activities (inputs) are evaluated.
  - **Organisation**: maintain an efficient production process (e.g. make sure the raw materials are sufficient and unerring on the line for workers to work with).
  - **Productivity**: increase the productivity of workers (e.g. keep the workers on track and focusing on the production task).
  - **Quality**: reduce line defect rates (e.g. constantly remind workers to use the standardised operating procedure in order to reduce the number of faulty products).
  - **Relationship**: team building (e.g. provide support and communication to foster a friendly and positive work environment).

The Treatment - Subjective Evaluation

- We used sliders instead of numbers to avoid ties and relative measures to alleviate measurement problems.





The Treatment - Tournament

- A tournament incentive determined by the aggregated results of the subjective evaluation in each month.
- Monthly rankings are independent.

	Initial Foreman Fee (RMB/M) (1)	Tournament Reward (RMB/M) (2)	Difference from the next lower rank (3)	Change in Total Foreman Fee (%) (4)
#.1 ranked forewoman	90	205	45	228%
#.2 ranked forewoman	90	160	25	178%
#.3 ranked forewoman	90	135	15	150%
#.4 ranked forewoman	90	120	10	133%
#.5 ranked forewoman	90	110	10	122%
#.6 ranked forewoman	90	100	10	111%
#.7 ranked forewoman	90	90		100%
Eliminated by the manager		0		

The Treatment - Tournament

- A tournament incentive determined by the aggregated results of the subjective evaluation in each month.
- Monthly rankings are independent.

	Initial Foreman Fee (RMB/M) (1)	Tournament Reward (RMB/M) (2)	Difference from the next lower rank (3)	Change in Total Foreman Fee (%) (4)
#.1 ranked forewoman	90	205	45	228%
#.2 ranked forewoman	90	160	25	178%
#.3 ranked forewoman	90	135	15	150%
#.4 ranked forewoman	90	120	10	133%
#.5 ranked forewoman	90	110	10	122%
#.6 ranked forewoman	90	100	10	111%
#.7 ranked forewoman	90	90		100%
Eliminated by the manager		0		

The Treatment - Tournament

- A tournament incentive determined by the aggregated results of the subjective evaluation in each month.
- Monthly rankings are independent.

	Initial Foreman Fee (RMB/M) (1)	Tournament Reward (RMB/M) (2)	Difference from the next lower rank (3)	Change in Total Foreman Fee (%) (4)
#.1 ranked forewoman	90	205	45	228%
#.2 ranked forewoman	90	160	25	178%
#.3 ranked forewoman	90	135	15	150%
#.4 ranked forewoman	90	120	10	133%
#.5 ranked forewoman	90	110	10	122%
#.6 ranked forewoman	90	100	10	111%
#.7 ranked forewoman	90	90		100%
Eliminated by the manager		0		

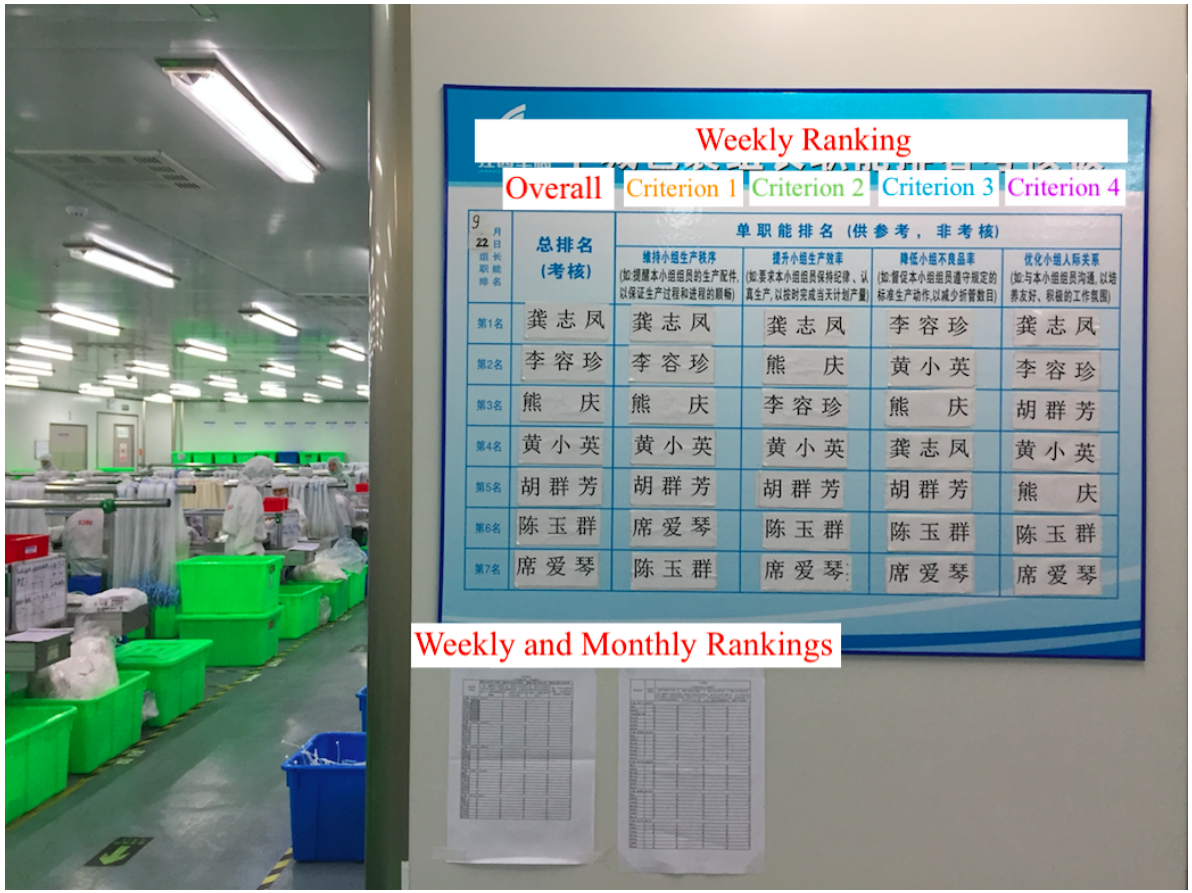
The Treatment - Tournament

- A tournament incentive determined by the aggregated results of the subjective evaluation in each month.
- Monthly rankings are independent.

	Initial Foreman Fee (RMB/M) (1)	Tournament Reward (RMB/M) (2)	Difference from the next lower rank (3)	Change in Total Foreman Fee (%) (4)
#.1 ranked forewoman	90	205	45	228%
#.2 ranked forewoman	90	160	25	178%
#.3 ranked forewoman	90	135	15	150%
#.4 ranked forewoman	90	120	10	133%
#.5 ranked forewoman	90	110	10	122%
#.6 ranked forewoman	90	100	10	111%
#.7 ranked forewoman	90	90		100%
Eliminated by the manager		0		

The Treatment - Relative Performance Feedback

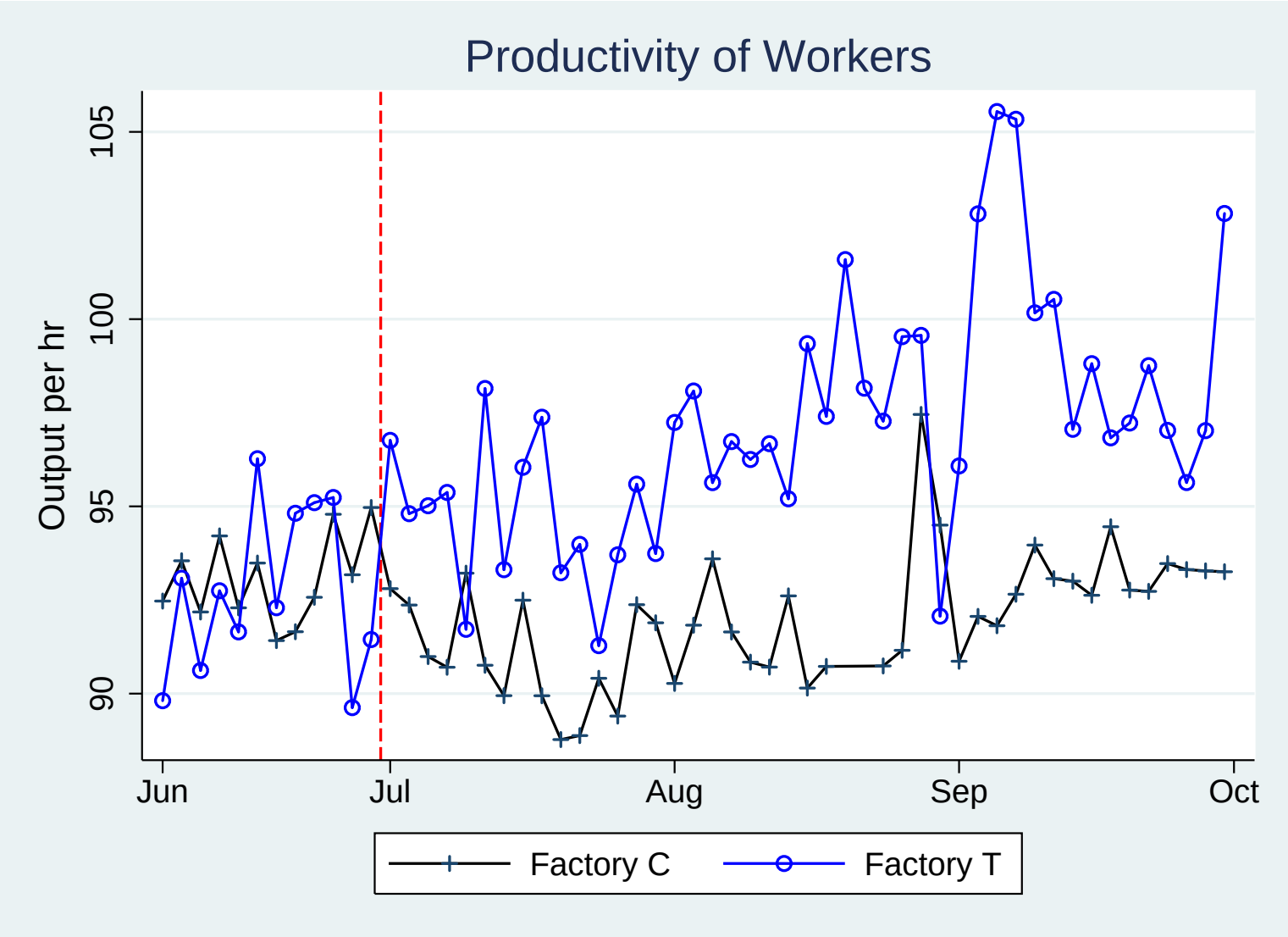
- To minimise favouritism and unfairness, rankings are publicly provided in factory T during the treatment period.



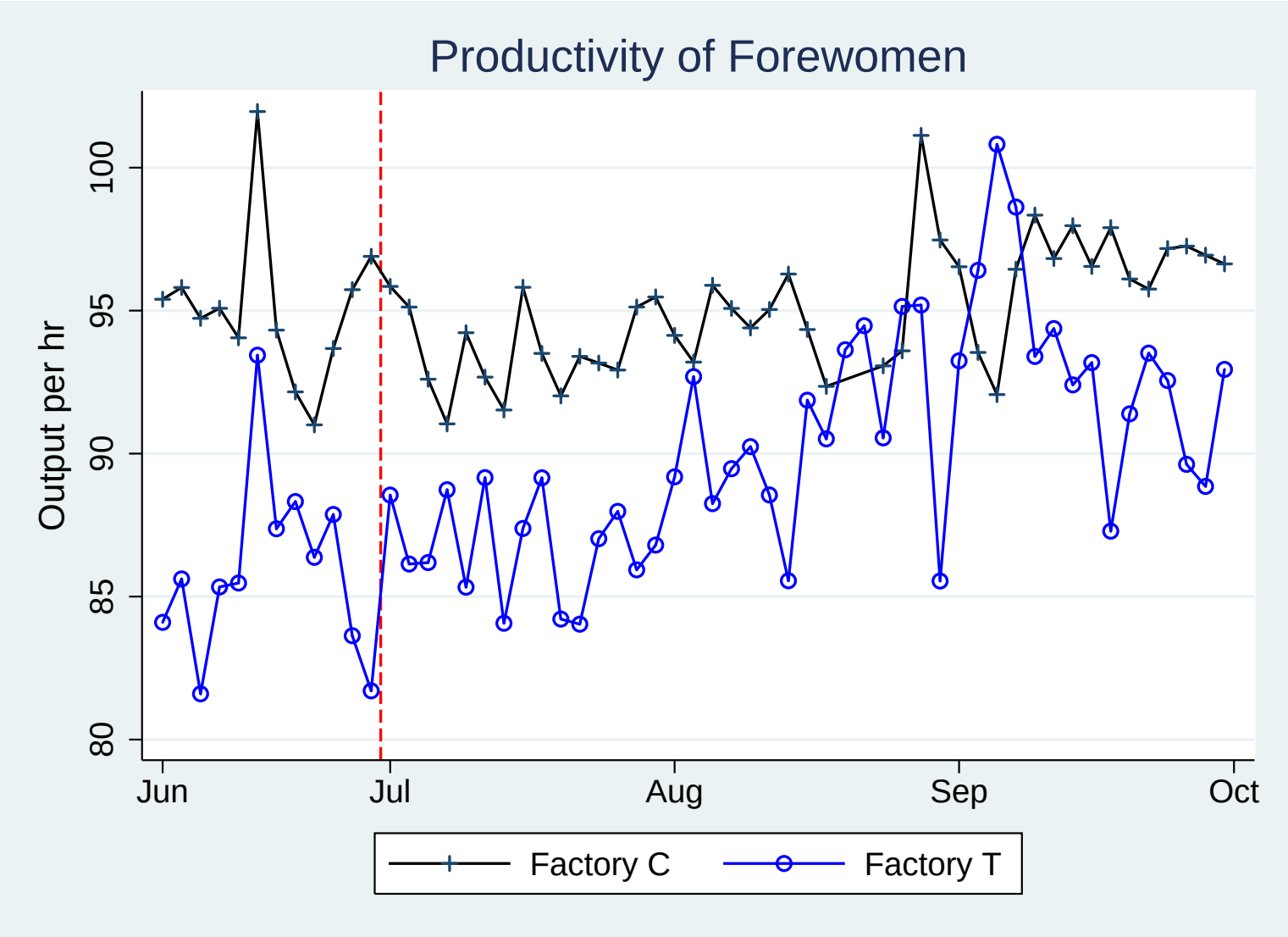
Hypotheses

- In factory T, workers' productivity increase as our treatment motivates forewomen to improve the team efficiency.
- For forewomen in factory T, twofold effects:
  - Positive: improved team efficiency increases productivity,
  - Negative: inputs to make others more productive reduce own productivity.

Results: Productivity of Workers Over Time



Results: Productivity of Forewomen Over Time



Main Results

- Graphs are in line with regression results accounting for individual, time fixed effects, and other time-varying factors.
  Table for workers, Table for forewomen,
  Parallel trend test for workers, Parallel trend test for forewomen.

- Compared to the control factory:
  - Incentivising forewomen on organisational activities increased the productivity of workers by 6%,
  - The treatment effect on the productivity of forewomen is positive but smaller and insignificant. Why?

- Forewomen increased working time, 30-50 mins per day in the first two months of treatment. Table for forewomen.

Main Results

- Graphs are in line with regression results accounting for individual, time fixed effects, and other time-varying factors.
  Table for workers, Table for forewomen,
  Parallel trend test for workers, Parallel trend test for forewomen.

- Compared to the control factory:
  - Incentivising forewomen on organisational activities increased the productivity of workers by 6%,
  - The treatment effect on the productivity of forewomen is positive but smaller and insignificant. Why?

- Forewomen increased working time, 30-50 mins per day in the first two months of treatment. Table for forewomen.

Other Results

- Workers also increased working time, but 15-20 mins less than the forewomen, but only weakly significant in the first month of treatment. Table for workers and DDD.
- Rankings vary over time: Monthly Rankings or Weekly Rankings.
- According to the management, soft management (e.g. team cohesion) also improved profoundly.
- Questionnaires indicate that both forewomen and workers were happy with the intervention.

Implications

- Our take: Subjective evaluation increased line managers' provision of a public good and increased team productivity.
- Profitable for the firm: overall cost of the treatment equals **half of one** worker's income, while increase in overall output equivalent to hiring **two** additional workers.
- The firm decided to keep using our design and try to apply it in other departments.

Table: The Treatment Effect on Worker's Performance

	Log(Output)				Log(Productivity (output per hour))			
	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.089***	-0.130***	-0.116***	-0.118***	-0.018	-0.016*	-0.015*	-0.041***
	(0.013)	(0.009)	(0.010)	(0.013)	(0.013)	(0.008)	(0.009)	(0.011)
Post	-0.038**	-0.060***	-0.090***	-0.017	0.024	0.000	0.030*	0.017
	(0.018)	(0.017)	(0.022)	(0.021)	(0.018)	(0.015)	(0.015)	(0.020)
Treatment*Post	0.079***	0.091***	0.096***	0.062***	0.061***	0.037***	0.053***	0.082***
	(0.017)	(0.016)	(0.019)	(0.023)	(0.016)	(0.014)	(0.016)	(0.020)
Observations	5,693	2,770	2,664	2,733	5,693	2,770	2,664	2,733
Clusters	57	57	57	57	57	57	57	57
R^2	0.530	0.659	0.493	0.593	0.754	0.782	0.742	0.763
Controls	YES	YES	YES	YES	YES	YES	YES	YES

Back

Table: The Treatment Effect on Forewoman's Production Performance

	Log(Output)				Log(Productivity (output per hour))			
	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.121***	-0.111***	-0.118***	-0.100***	-0.101***	-0.099***	-0.096***	-0.090***
	(0.024)	(0.014)	(0.023)	(0.023)	(0.016)	(0.012)	(0.012)	(0.015)
Post	-0.056*	-0.088**	-0.121***	-0.029	0.027	-0.003	0.030	0.030
	(0.031)	(0.039)	(0.039)	(0.032)	(0.026)	(0.030)	(0.027)	(0.028)
Treatment*Post	0.082**	0.100***	0.099**	0.060	0.034	0.023	0.042	0.039
	(0.031)	(0.028)	(0.040)	(0.042)	(0.021)	(0.023)	(0.023)	(0.025)
Observations	1,312	644	621	621	1,312	644	621	621
Clusters	13	13	13	13	13	13	13	13
R^2	0.350	0.491	0.314	0.497	0.845	0.873	0.832	0.840
Controls	YES	YES	YES	YES	YES	YES	YES	YES

Back

Table: The Treatment Effect on Worker's Working Time

	Number of Minutes Worked in a Day			
	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep
	(1)	(2)	(3)	(4)
Treatment	-50.999***	-75.000***	-63.183***	-55.082***
	(7.231)	(5.305)	(5.376)	(6.688)
Post	-35.510***	-34.317***	-48.690***	-22.542**
	(10.253)	(9.792)	(10.481)	(10.794)
Treatment*Post	8.345	37.247***	11.183	-9.921
	(9.412)	(10.049)	(9.920)	(11.848)
Observations	5,693	2,770	2,664	2,733
Clusters	57	57	57	57
R^2	0.410	0.572	0.404	0.482
Controls	YES	YES	YES	YES

Back

Table: The Treatment Effect on Forewoman's Working Time

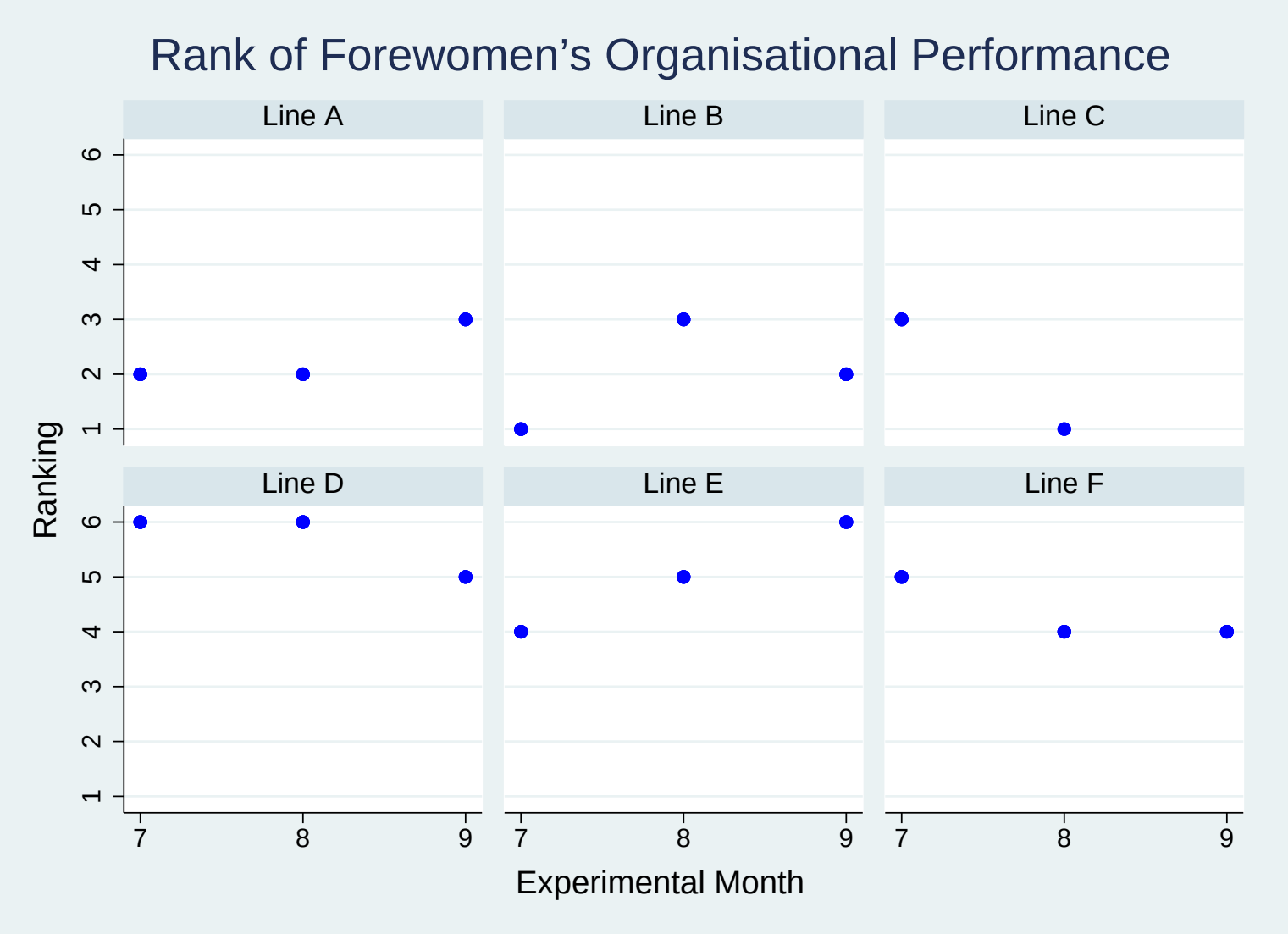
	Number of Minutes Worked in a Day			
	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep
	(1)	(2)	(3)	(4)
Treatment	-17.348	-16.302	-19.074	-5.901
	(18.453)	(10.845)	(16.900)	(17.374)
Post	-48.913***	-46.208*	-60.541***	-33.305*
	(14.022)	(22.191)	(15.984)	(15.781)
Treatment*Post	24.024	54.378**	28.904	0.293
	(23.912)	(19.726)	(30.201)	(32.761)
Observations	1,312	644	621	621
Clusters	13	13	13	13
R^2	0.314	0.416	0.293	0.353
Controls	YES	YES	YES	YES

Back

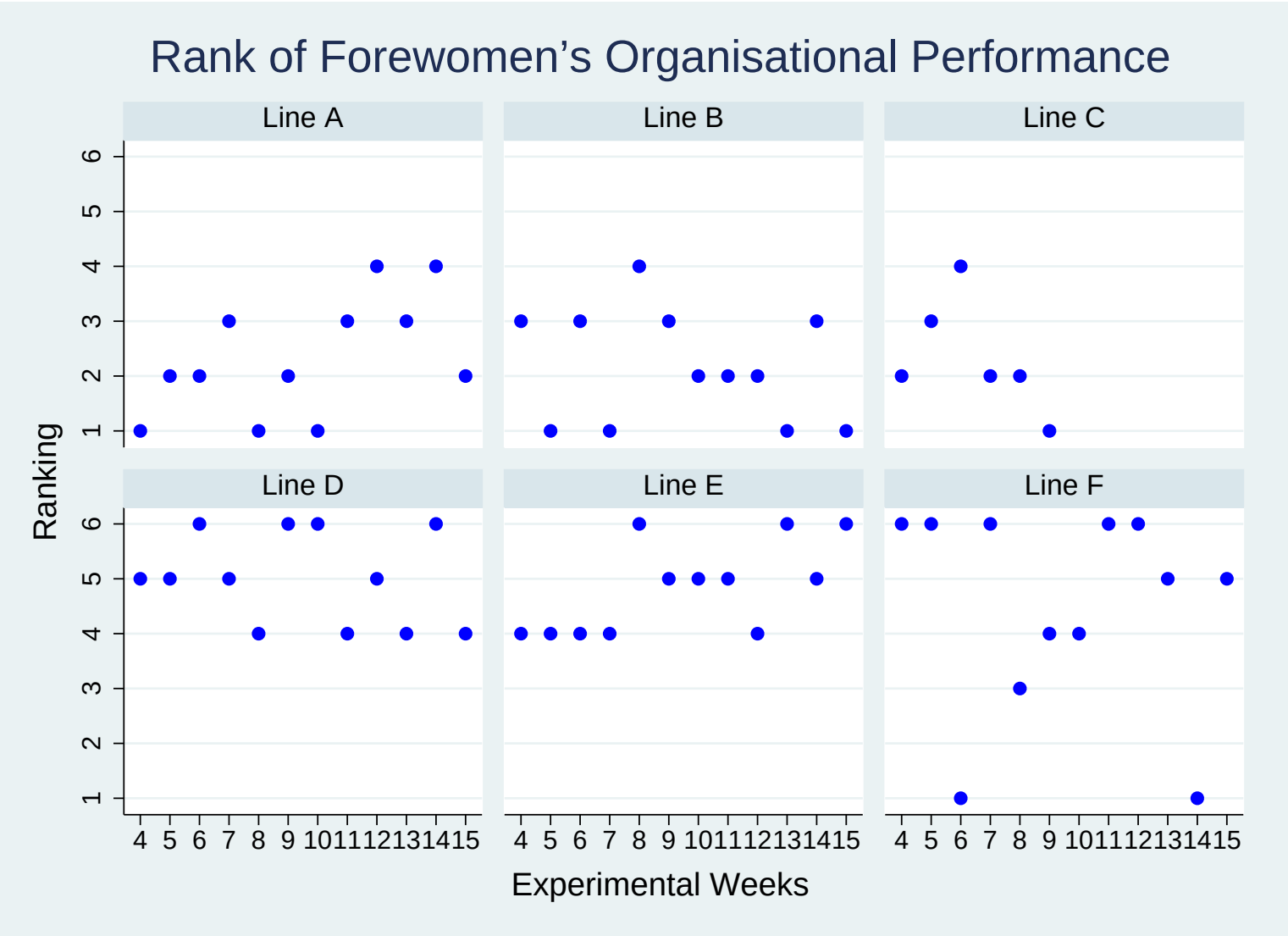
Table: DDD of the Treatment Effect on Forewoman's Working Time Comparing to Workers

	Number of Minutes Worked in a Day			
	Jun vs Jul-Sep	Jun vs Jul	Jun vs Aug	Jun vs Sep
	(1)	(2)	(3)	(4)
Treatment	-51.015***	-75.074***	-63.161***	-54.995***
	(13.482)	(19.344)	(18.029)	(18.944)
Post	-36.811	-34.936	-50.555	-23.433
	(23.655)	(27.960)	(45.984)	(23.899)
Foreman	-25.754***	-32.156**	-25.878***	-20.035**
	(7.008)	(12.802)	(7.801)	(7.726)
Treatment*Post	8.688	37.037*	11.593	-9.717
	(16.002)	(20.296)	(33.506)	(16.591)
Foreman*Post	-6.848	-8.394	-0.435	-6.087
	(5.106)	(7.304)	(5.579)	(5.634)
Foreman*Treatment	14.442	38.232*	18.301	-6.071
	(12.445)	(19.873)	(15.380)	(17.158)
Foreman*Treatment*Post	14.120	18.827*	14.704	9.274
	(9.282)	(11.088)	(10.563)	(12.505)
Observations	7,005	3,414	3,285	3,354
Clusters	215	102	101	104
R^2	0.390	0.542	0.380	0.457
Controls	YES	YES	YES	YES

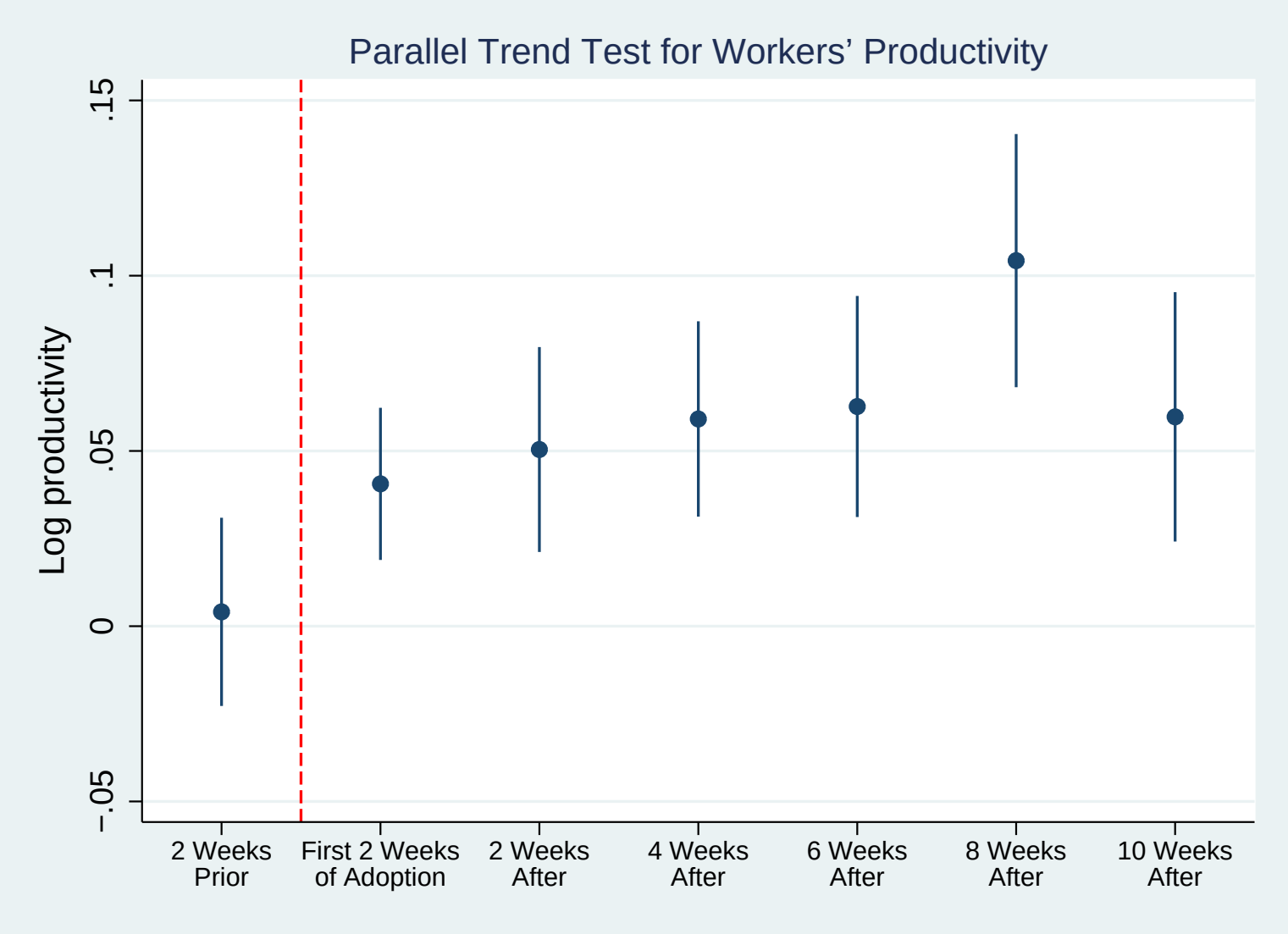
Back


Rank of Forewomen's Organisational Performance
Line A
Line B
Line C
Line D
Line E
Line F
Ranking
1
2
3
4
5
6
7
8
9
Experimental Month

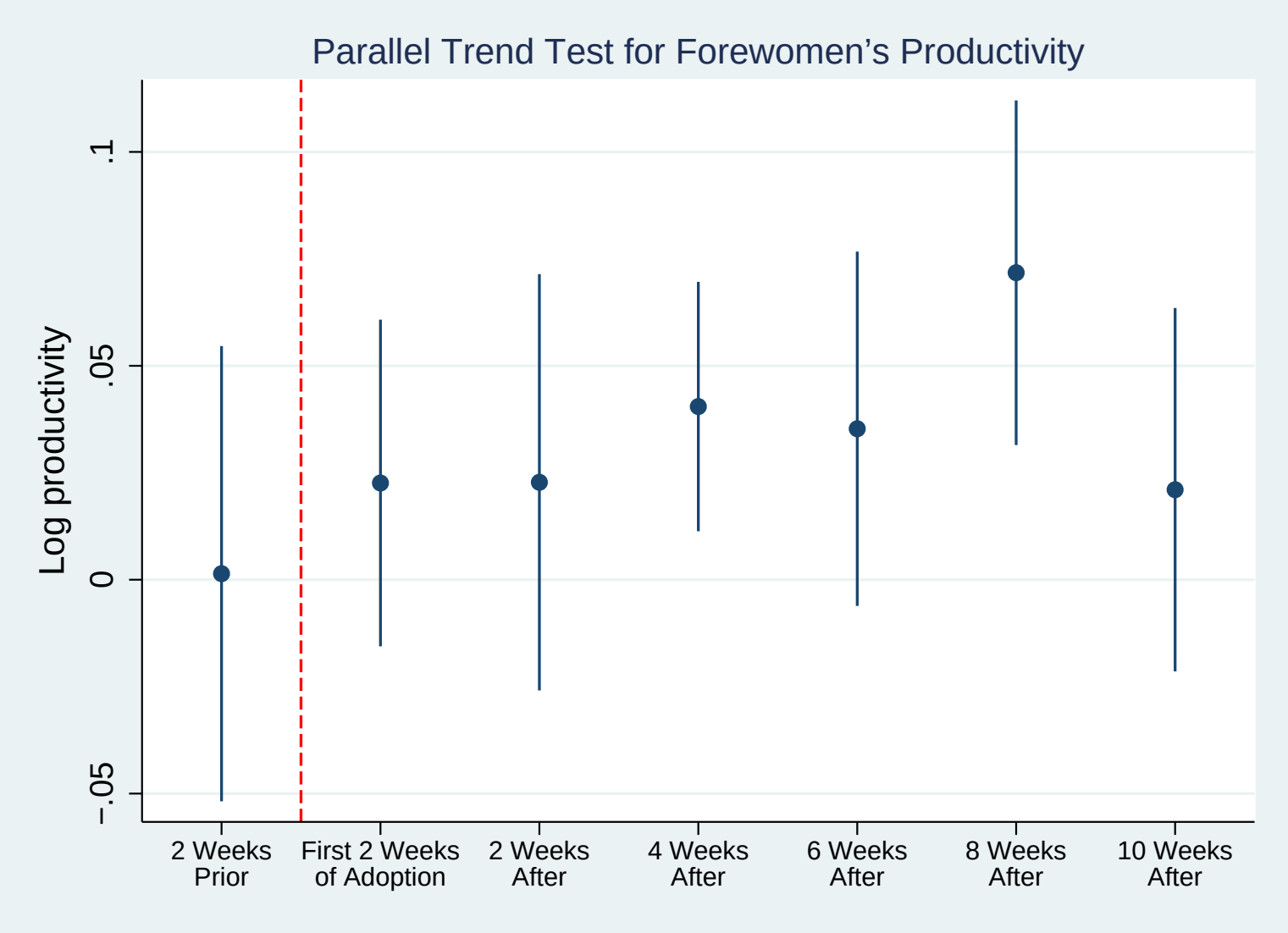
Back


Rank of Forewomen's Organisational Performance
Line A
Line B
Line C
Line D
Line E
Line F
Ranking
1 2 3 4 5 6
4 5 6 7 8 9 10 11 12 13 14 15
Experimental Weeks

Back


Parallel Trend Test for Workers' Productivity
Log productivity
−.05
0
.05
.1
.15
2 Weeks Prior
First 2 Weeks of Adoption
2 Weeks After
4 Weeks After
6 Weeks After
8 Weeks After
10 Weeks After

Back


Parallel Trend Test for Forewomen's Productivity
Log productivity
−.05
0
.05
.1
2 Weeks Prior
First 2 Weeks of Adoption
2 Weeks After
4 Weeks After
6 Weeks After
8 Weeks After
10 Weeks After

Back