July 2024

Productive Creators: A New Studio Indicator

ROBLOX

The Role of Indicators

Studio Engineering → Indicators → Outcomes

(e.g. Gen AI Tool) (e.g. Retention) (e.g. Playtime)

What is a good indicator?

- Strong causal driver of business outcomes.
- Movable— connected to creator behaviours that can be nurtured.
- Predictable.
- Easy to build.
- Easy to communicate.

Studio Indicators

Indicator	Source Data	Size[1]	Definition	Concern
Total Creators	Time Spent	9.4M	Total users who downloaded Studio.	Too Broad
Active (L28)	Time Spent	3.7M	No. of users with time spent > 0.	Not quality adjusted.
Time Spent (L28)	Time Spent	392K Hours	Time spent by anyone with an active studio session.	Not quality adjusted.
New User Retention (L7)	Time Spent	60%	% active in current period, over those active in current & past.	Not quality adjusted.
Funnel (Playtime): Inactive, Active, Core, Top	Time Spent + Playtime	5.7M, 3.6M, 50K, 2.3K	Number of active creators with 0, 0-100, 100-100K, 100K+ amount of playtime in hours.	• Is actually an outcome • Not an intrinsic behaviour • Does not rule out luck
★ Productive Creators (L28)	Time Spent + Collaboration + Publishing +Tenure	0.42M	(Long Tenure OR Recently Engaged) AND (Collaborating OR Publishing)	Complex, involves multiple series and thresholds.

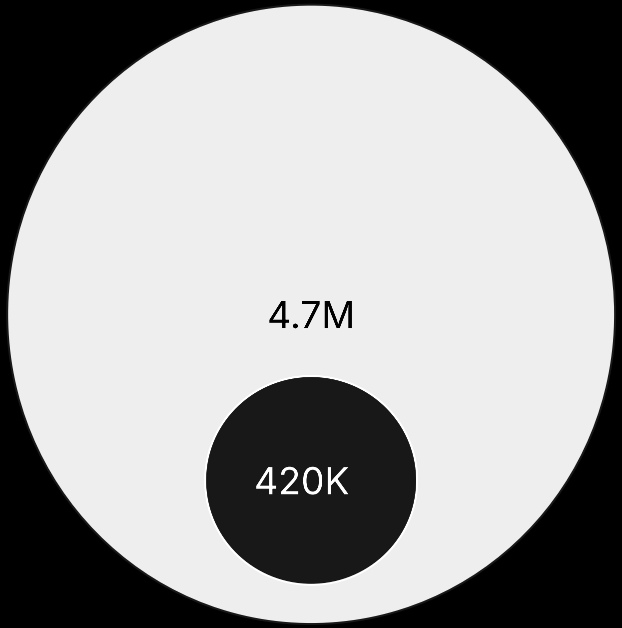
[1]Data is taken from dw_nonpii.developer_creator_velocity on ds = '2024-07-01'. Query link.

Productive Creators: Definition

Only <10% of active creators also productive.

Productivity:
(1000H Tenure OR Active L28>4)
AND
(Collaboration L28 >=1 OR Team Publishing>3)

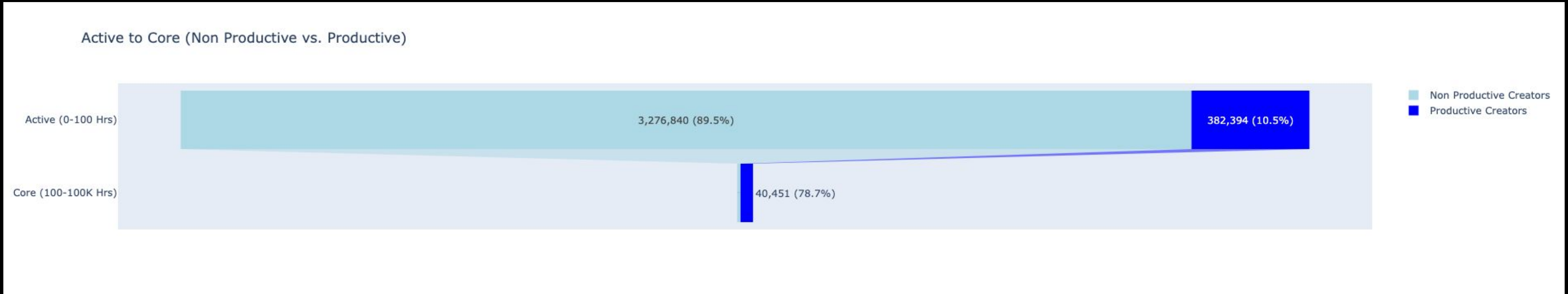
“Quantity of Time x Quality of Time”

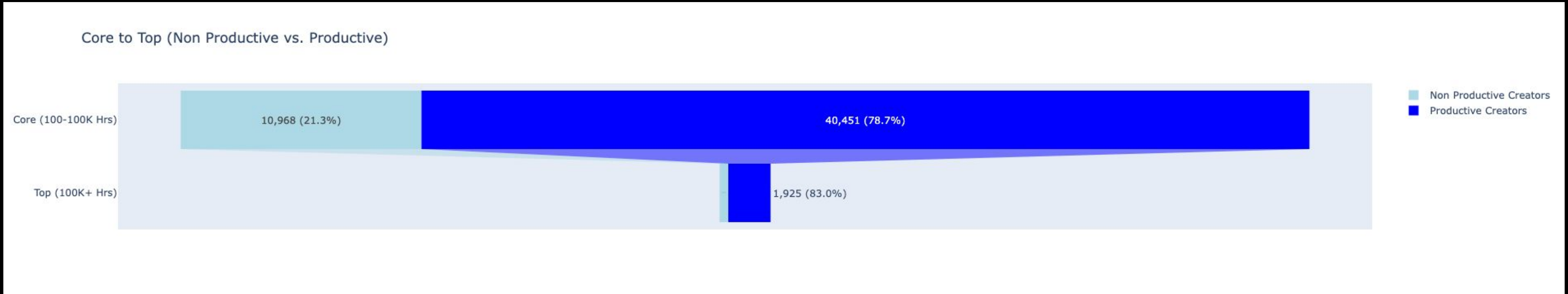


Productive in Active (L28) in 2024-07-01

Funnel

About ~80% of current top and core creators are productive.





ROBLOX

[1]Data is taken from dw_nonpii.developer_creator_velocity on ds = '2024-07-01'. Query link.

Studio Eng → Productivity → Playtime

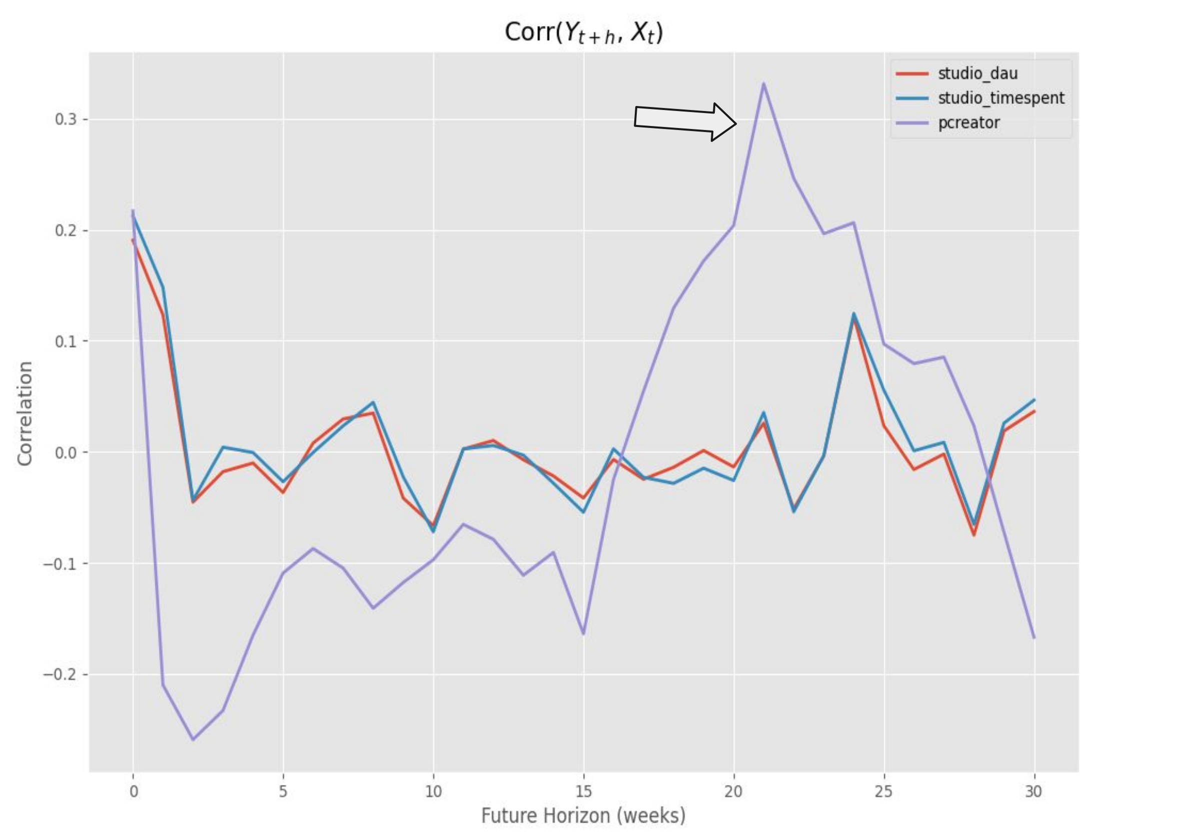
Platform Level Correlations

Y: Agg. Playtime WoW Growth
X: Agg. Indicator WoW Growth
Period: 2022-10 to 2024-07

Productive Creators correlated to *future* playtime.

Problems:
- Y(t-1), X(t-1) → Y(t+h)
- Y(t-1), X(t-1) → X(t)
- How far back?

Solution:
Control for as many lags without overfitting.



ROBLOX

Platform Level VAR

Vector AutoRegression:

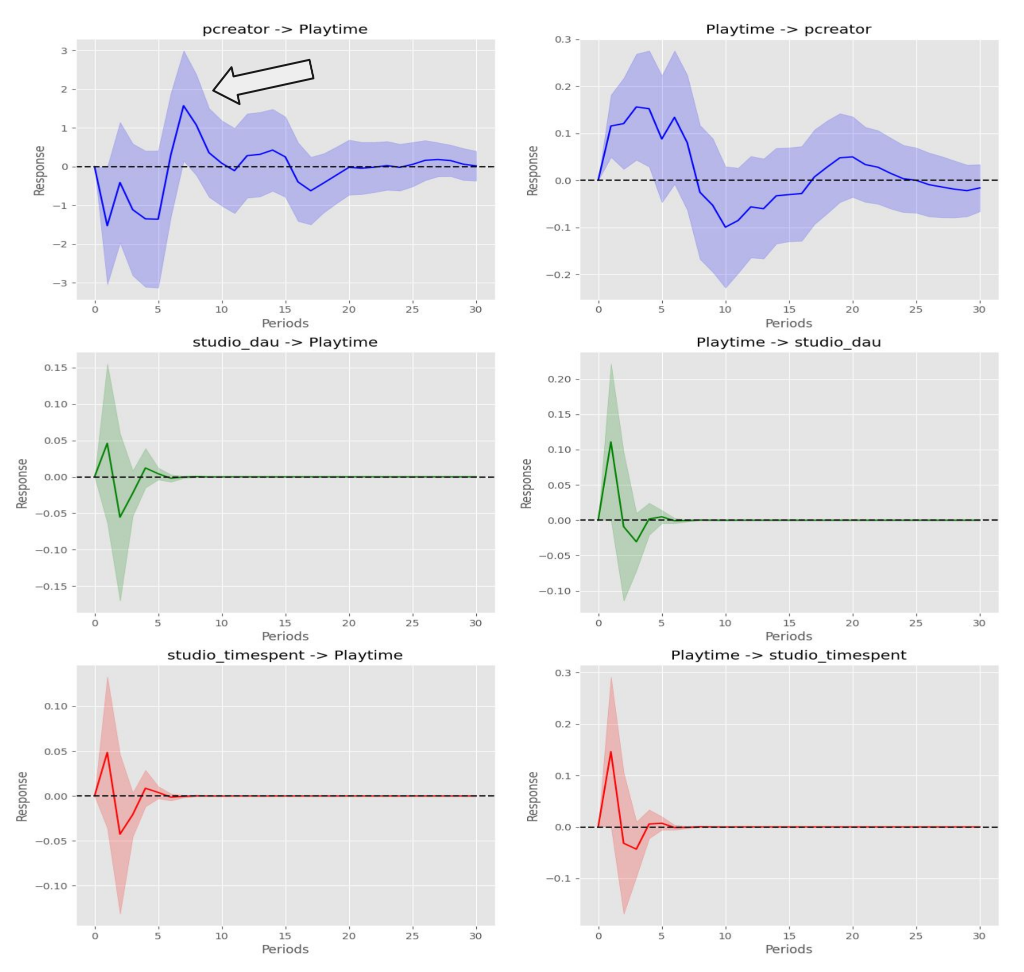
$$Y_t = \alpha^Y + \sum_r \beta_r^Y Y_{t-r} + \sum_r \tau_r^Y X_{t-r} + \epsilon_t^Y$$

$$X_t = \alpha^X + \sum_r \beta_r^X X_{t-r} + \sum_r \tau_r^Y Y_{t-r} + \epsilon_t^X$$

A 1 pp jump in productivity growth leads to a 1.5 pp jump in playtime growth in 8 weeks.

Indicator	Lag Order Chosen
DAU	2
Time Spent	1
Productive	9

Impulse Responses



Tests of Predictability

- Granger Causality Test: Can series X predict *future* values of series Y?
  - Null Hypothesis: All Taus are 0.

$$Y_t = \alpha^Y + \sum_r \beta_r^Y Y_{t-r} + \sum_r \tau_r^Y X_{t-r} + \epsilon_t^Y$$

$$X_t = \alpha^X + \sum_r \beta_r^X X_{t-r} + \sum_r \tau_r^Y Y_{t-r} + \epsilon_t^X$$

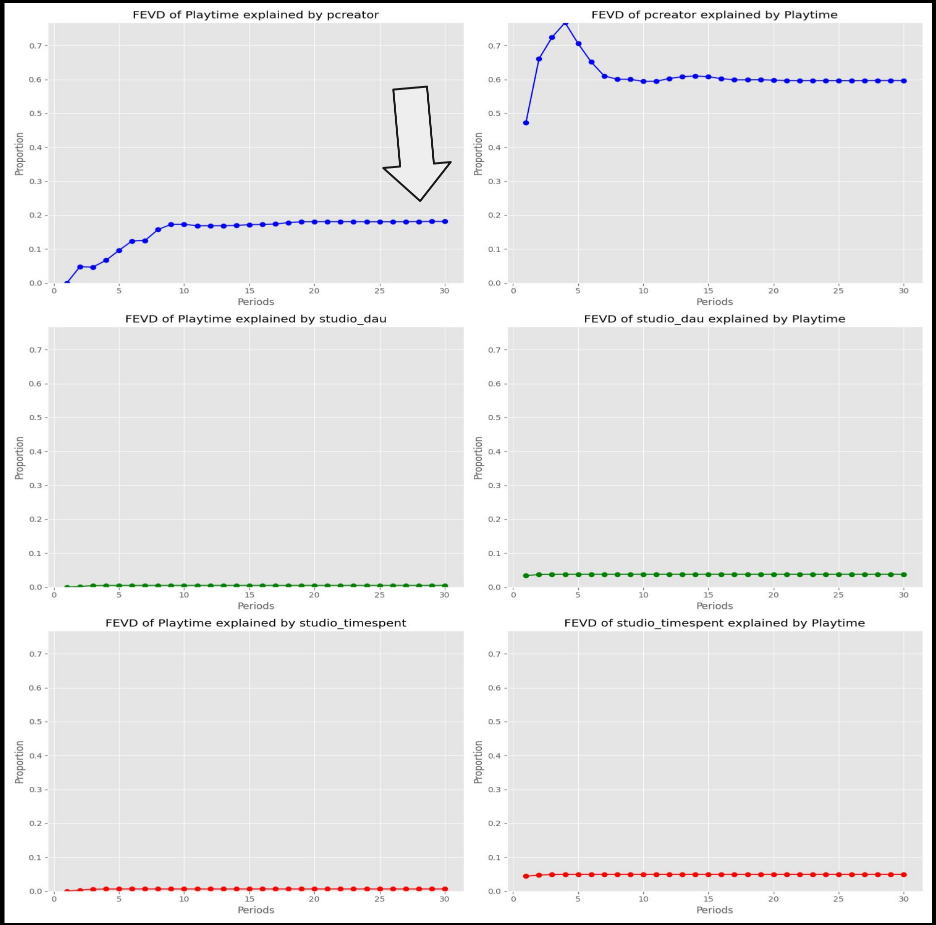
Indicator	Indicator → Playtime (P-Value)	Playtime → Indicator (P-Value)
DAU	0.282	0.1350
Time Spent	0.177	0.1304
Productive	0.044	0.001

Only playtime and productivity predict each other.

 Full set of results: here.

Forecast Error Variance Decomposition

How much one series is explained by shocks to the other series?

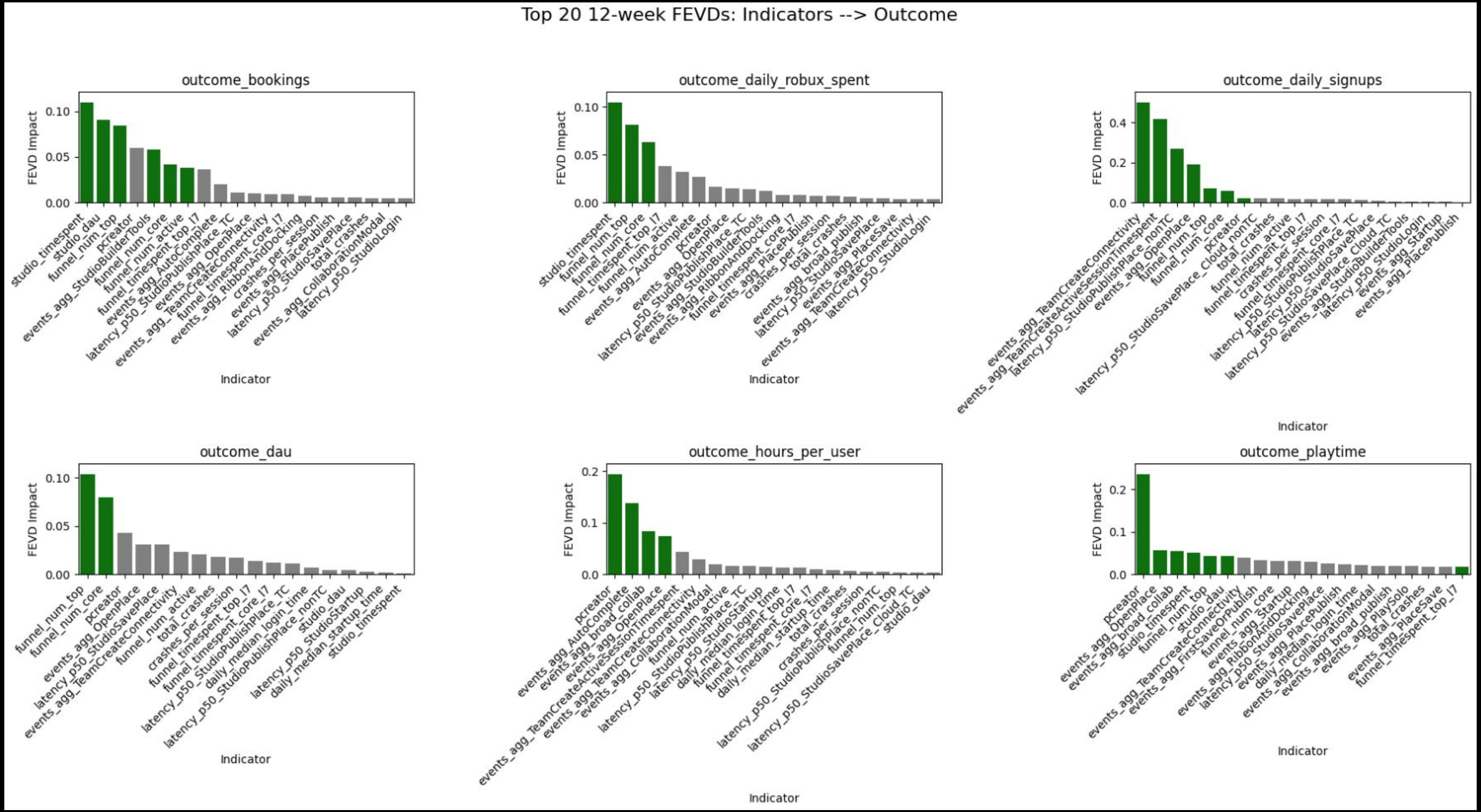


Productivity growth explains 20% of future playtime growth.

Playtime growth explains 60% of productivity growth.

Full set of results: here.

Extended Variance Decomposition



ROBLOX

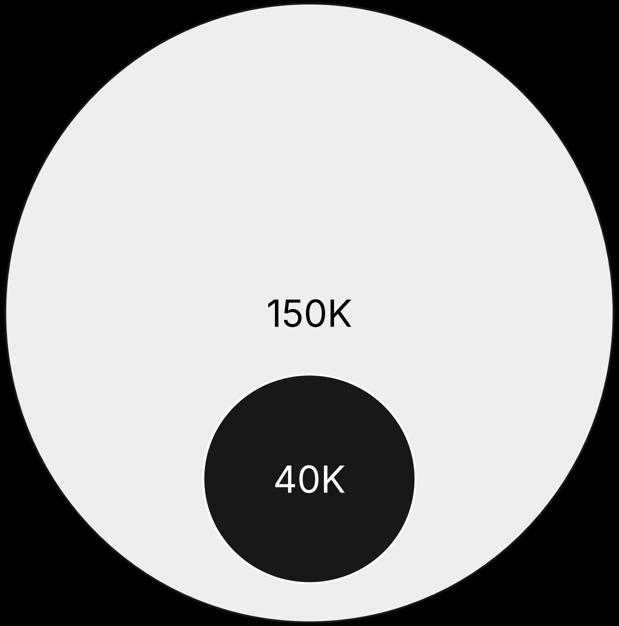
Total of 135 indicators tested including funnel, crash/latency, raw and agg events, etc.
Highlighted are those that passed GC Test of Indicator → Outcome. Lag orders chosen by AIC.

What if a creator was productive for a long time?

How much does it matter compared to mere activity?

Consistently Productive: Being a Productive Creator on the first date of every month, for 9 months.

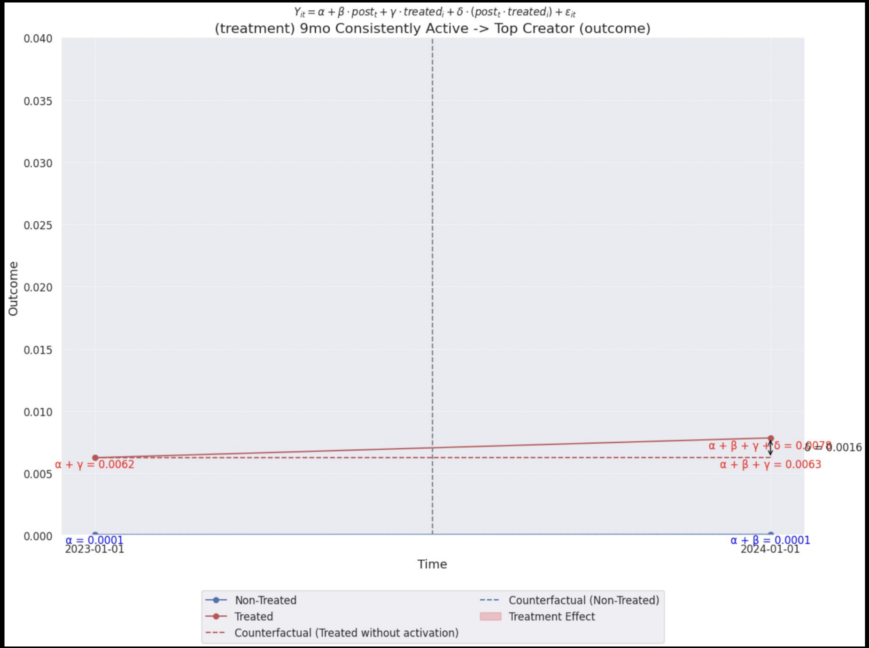
Consistently Active: Being a Active L28 at least once on first date of every month, for 9 months.

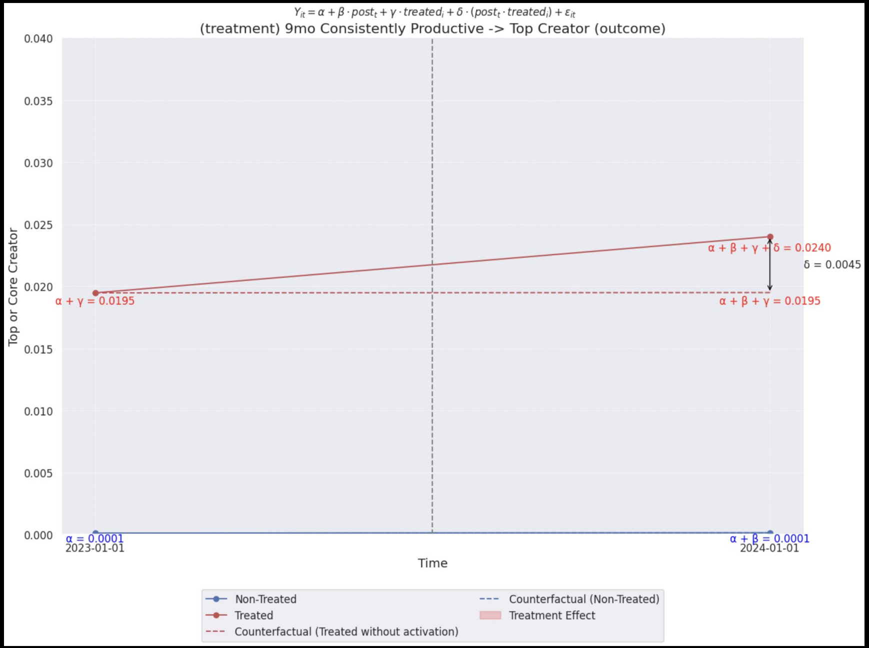


Consistently Productive vs Active between 2023-02 to 2023-10 (9 months).

Outcome: Prob(Top Creator in Future)

The impact of being "productive" on Prob(Top in Future) is ~280% more than simply being active.

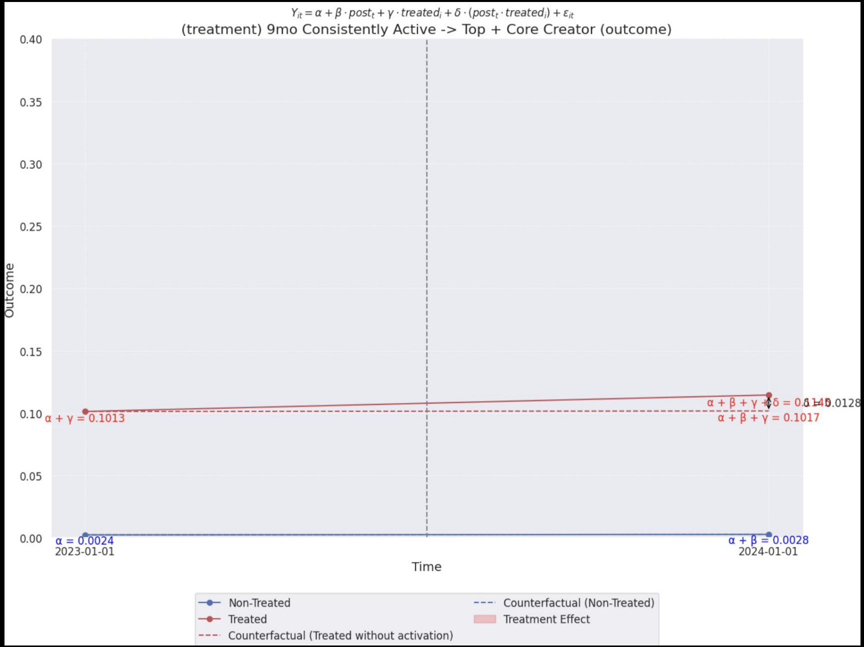


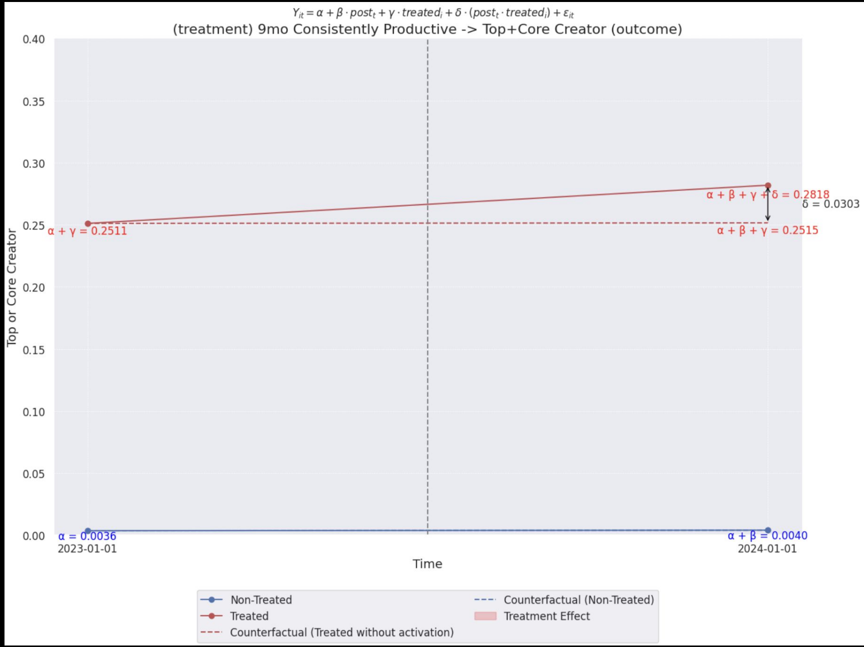


Consistent productivity/activity captured at end of each month from 2023-02 to 2023-10. For outcomes, Pre Period: 2023-01 and Post Period: 2024-01.

Outcome: Prob(Top or Core Creator in Future)

The impact of being “productive” on Prob(Top or Core in Future) is ~240% more than simply being active.

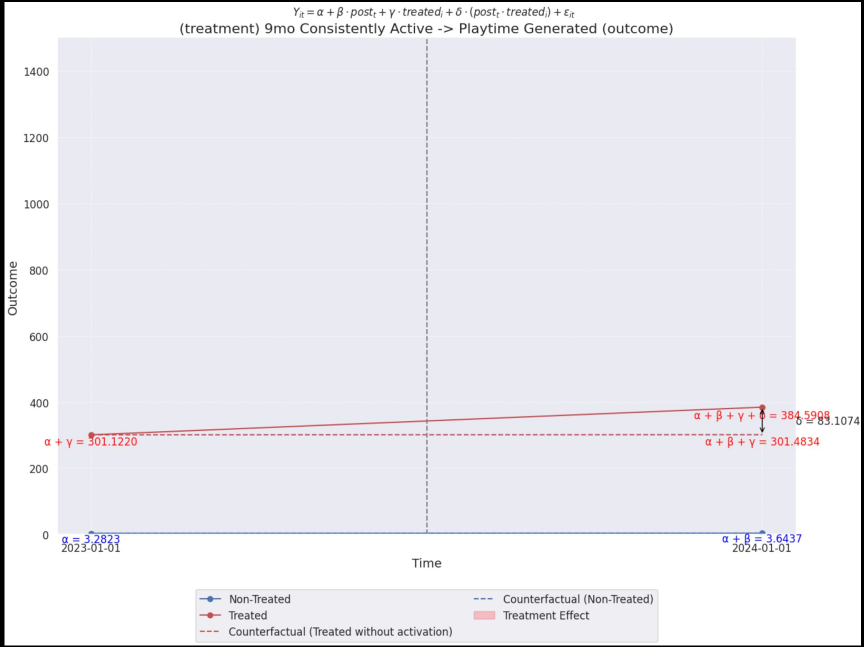


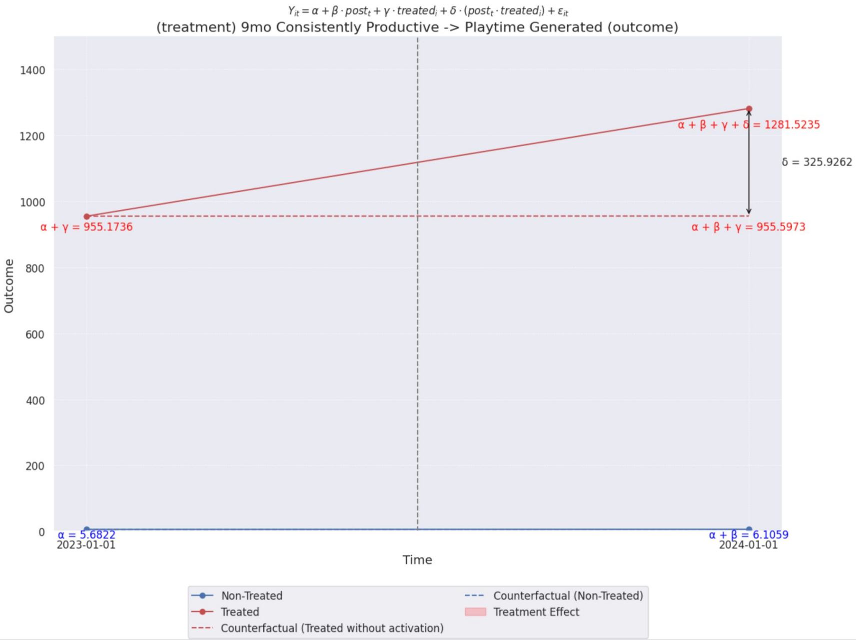


Consistent productivity/activity captured at end of each month from 2023-02 to 2023-10. For outcomes, Pre Period: 2023-01 and Post Period: 2024-01.

Active/Productive → Playtime Generated in Future

The impact of being "productive" on Playtime Generated is ~400% more than simply being active.





Consistent productivity/activity captured at end of each month from 2023-02 to 2023-10. For outcomes, Pre Period: 2023-01 and Post Period: 2024-01.

Lift

1.5 pp Lift to Playtime Growth (Weekly)

1% pp inc in Productive Creators

35% Lift to Playtime Generated (Annual)

Creator is Consistently Productive

Studio Eng → Productivity → Playtime

Onboarding Tutorial → Productivity

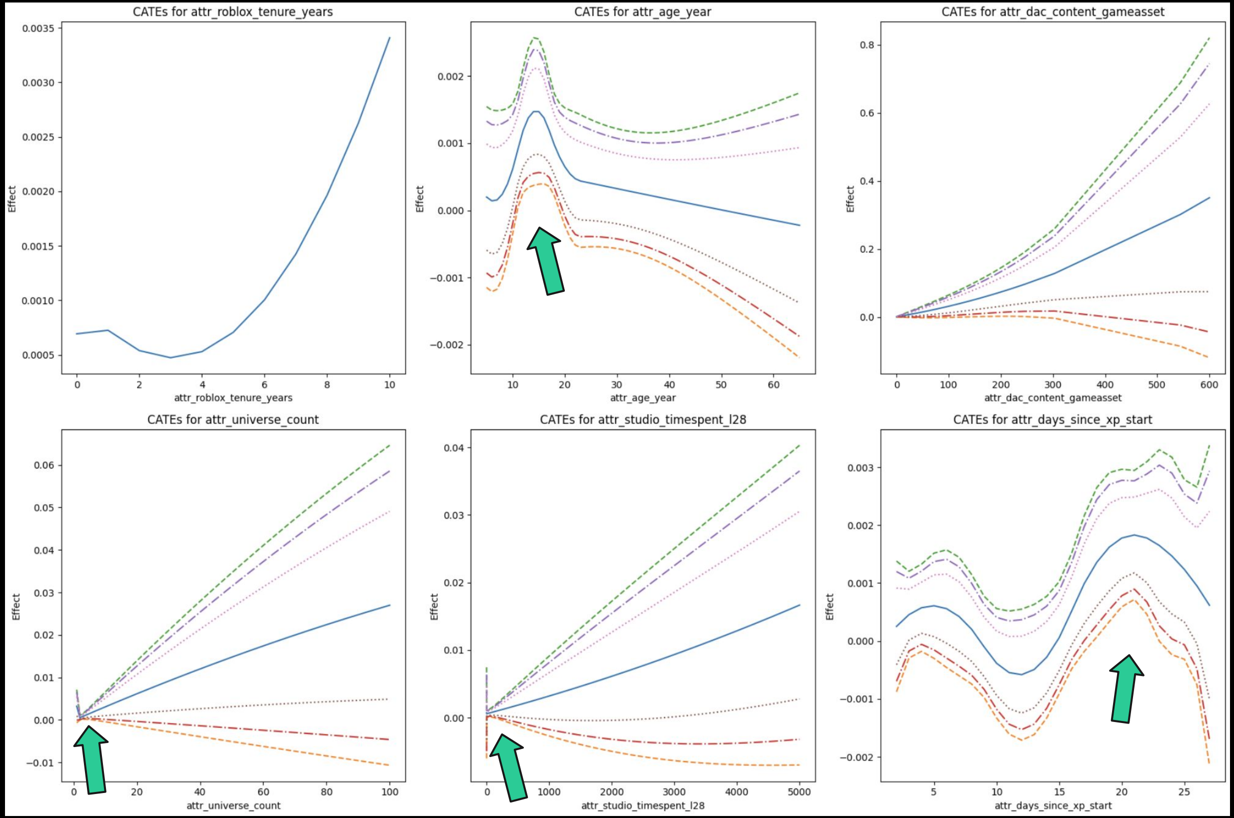
Pos stat sig impact of new studio tutorial on productivity of new creators.

Estimand	Model	Equations	Estimate (b)	Std. Error	t-value	p-value
ATE for Enrollment	Bivariate OLS	Y = a + b*Z + e	0.0007	0.0003	2.5536	0.0107
ATE for Enrollment	Multivariate OLS	Y = a + b*Z +c*D + e	0.0007	0.0003	2.4406	0.0147
ATE for Enrollment	DML	Y = b*Z +g(X) + e	0.0007	0.0003	2.386	0.017
ATE for Enrollment	DML (Interactive)	Y = g(Z,X) + e	0.0007	0.0003	2.5536	0.0107
LATE for Compliance	Bivariate IV	Y = a + b*T + e T = c + d*Z + u	0.0025	0.001	2.5545	0.0106
LATE for Compliance	Multivariate IV	Y = a + b*T + c*X + e T = e + f*Z + g*X + u	0.0024	0.001	2.477	0.0132
LATE for Compliance	DML IV	Y = b*T + g(X) + e T = m(Z,X) + u	0.0023	0.001	2.3701	0.0178
LATE for Compliance	DML IV (Interactive)	Y = g(T,X) + e T = m(Z,X) + u	0.0033	0.001	3.2118	0.0013

Y:Productive/Not, T:Compliance, Z:Enrollment, X: Covariates. N=1.4 million, First stage: AUC of T on Z is 0.8, on T+X is 0.85. Enrollment rate 50%, compliance rate 14%. Launch date: 08-2023 for 25 days. Eligibility is joining 1 month prior to experiment.

CATE

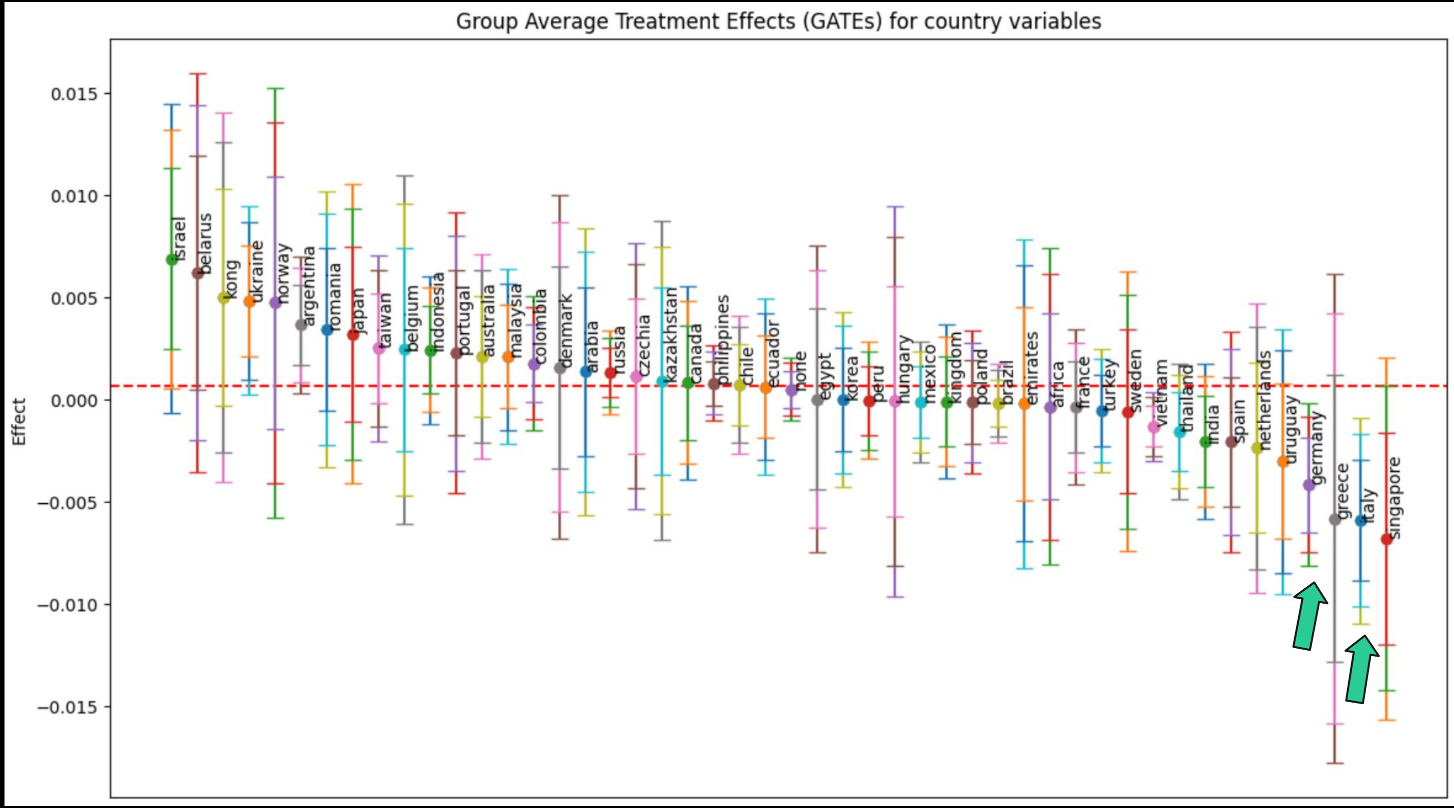
Stat sig CATE:
- In 15 year olds
- Early enrollers
- With 2-4 universes.
- 0-50m Time Spent



Method: DML-IRM

GATE

Tutorial failed to work in Italy, Germany.



Code Assist (GPT)→ Productivity

No stat sig impact of code assist (GPT) on productivity of scripters.

Estimand	Model	Equations	Estimate (b)	Std. Error	t-value	p-value
ATE for Enrollment	Bivariate OLS	Y = a + b*Z + e	0.0058	0.0048	1.1902	0.234
ATE for Enrollment	Multivariate OLS	Y = a + b*Z +c*D + e	0.0047	0.0045	1.0403	0.2982
ATE for Enrollment	DML	Y = b*Z +g(X) + e	0.0056	0.0042	1.3322	0.1828
ATE for Enrollment	DML (Interactive)	Y = g(Z,X) + e	0.0063	0.0043	1.4663	0.1426
LATE for Compliance	Bivariate IV	Y = a + b*T + e T = c + d*Z + u	1.1459	1.1216	1.0217	0.3069
LATE for Compliance	Multivariate IV	Y = a + b*T + c*X + e T = e + f*Z + g*X + u	1.004	1.1216	0.8952	0.3707
LATE for Compliance	DML IV	Y = b*T + g(X) + e T = m(Z,X) + u	1.4453	1.4521	0.9953	0.3196
LATE for Compliance	DML IV (Interactive)	Y = g(T,X) + e T = m(Z,X) + u	1.3256	1.2582	1.0536	0.2921

Y: Productive/Not, T: Compliance, Z: Enrollment, X: Covariates. N=85k;
Feb 2024 A/B Test: Control: 20%, GPT: 40%, CodeLlama: 40%.
Weak instrument problem.

Code Assist (CodeLlama)→ Productivity

No stat sig impact of code assist (CodeLlama) on productivity of scripters.

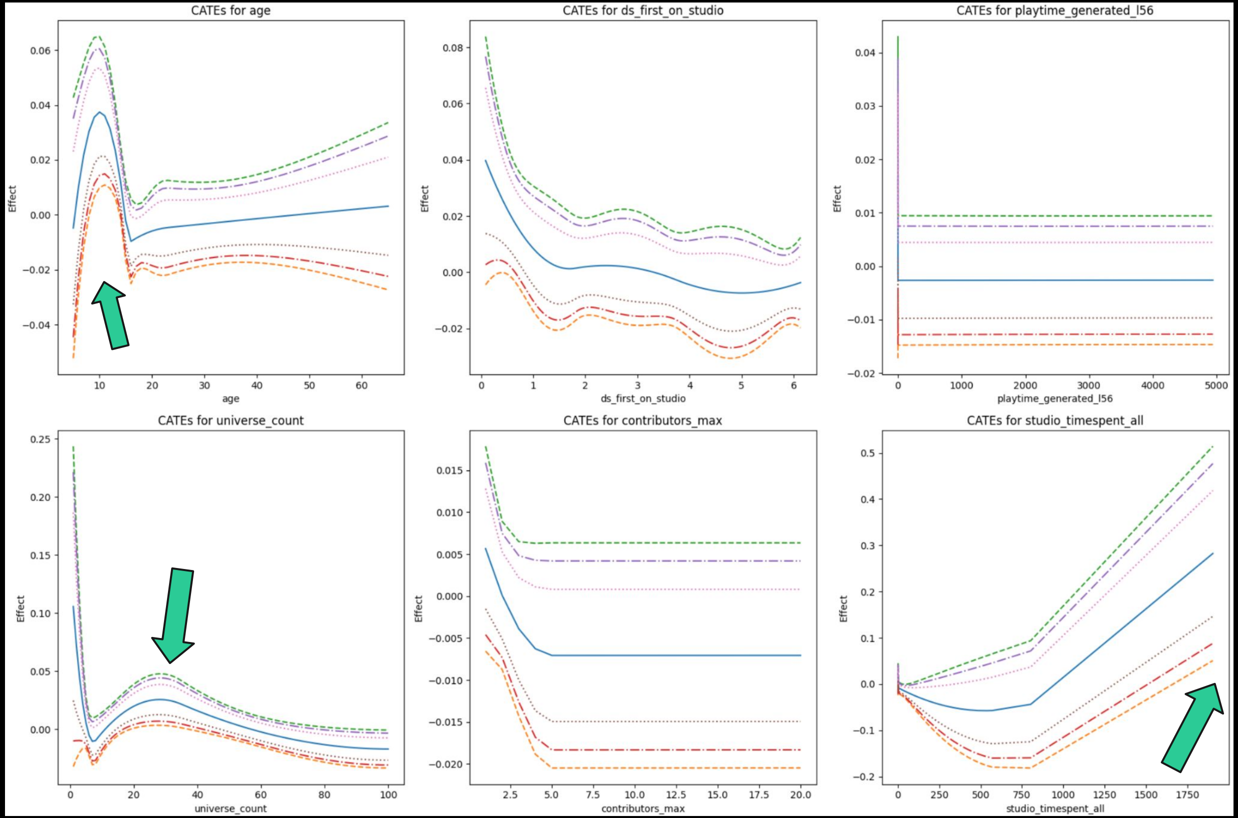
Estimand	Model	Equations	Estimate (b)	Std. Error	t-value	p-value
ATE for Enrollment	Bivariate OLS	Y = a + b*Z + e	0.0005	0.0048	0.104	0.9172
ATE for Enrollment	Multivariate OLS	Y = a + b*Z +c*D + e	0.0000	0.0045	-0.0087	0.993
ATE for Enrollment	DML	Y = b*Z +g(X) + e	-0.0003	0.0042	-0.0771	0.9386
ATE for Enrollment	DML (Interactive)	Y = g(Z,X) + e	-0.0004	0.0043	-0.0924	0.9264
LATE for Compliance	Bivariate IV	Y = a + b*T + e T = c + d*Z + u	0.1019	0.976	0.1044	0.9168
LATE for Compliance	Multivariate IV	Y = a + b*T + c*X + e T = e + f*Z + g*X + u	-0.0085	0.9732	-0.0087	0.993
LATE for Compliance	DML IV	Y = b*T + g(X) + e T = m(Z,X) + u	-0.0605	0.9155	-0.066	0.9473
LATE for Compliance	DML IV (Interactive)	Y = g(T,X) + e T = m(Z,X) + u	-0.0667	0.9183	-0.0727	0.9421

Y: Productive/Not, T: Compliance, Z: Enrollment, X: Covariates. N=85k;
Feb 2024 A/B Test: Control: 20%, GPT: 40%, CodeLlama: 40%.
Weak instrument problem.

CodeLlama: CATE

Stat sig CATE:

- 10 year olds
- 0-50 minutes spent on studio
- 20-60 universes.

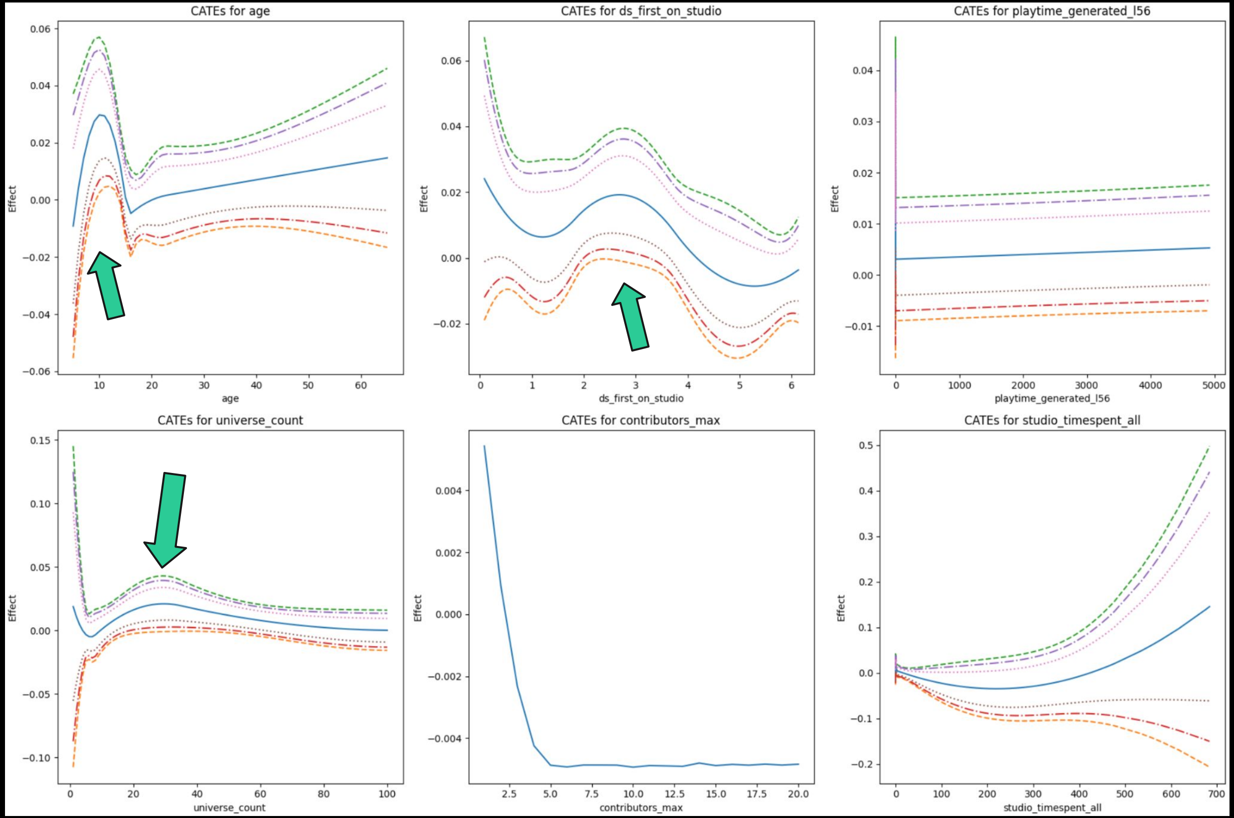


Method: DML-IRM

GPT: CATE

Pos stat sig:

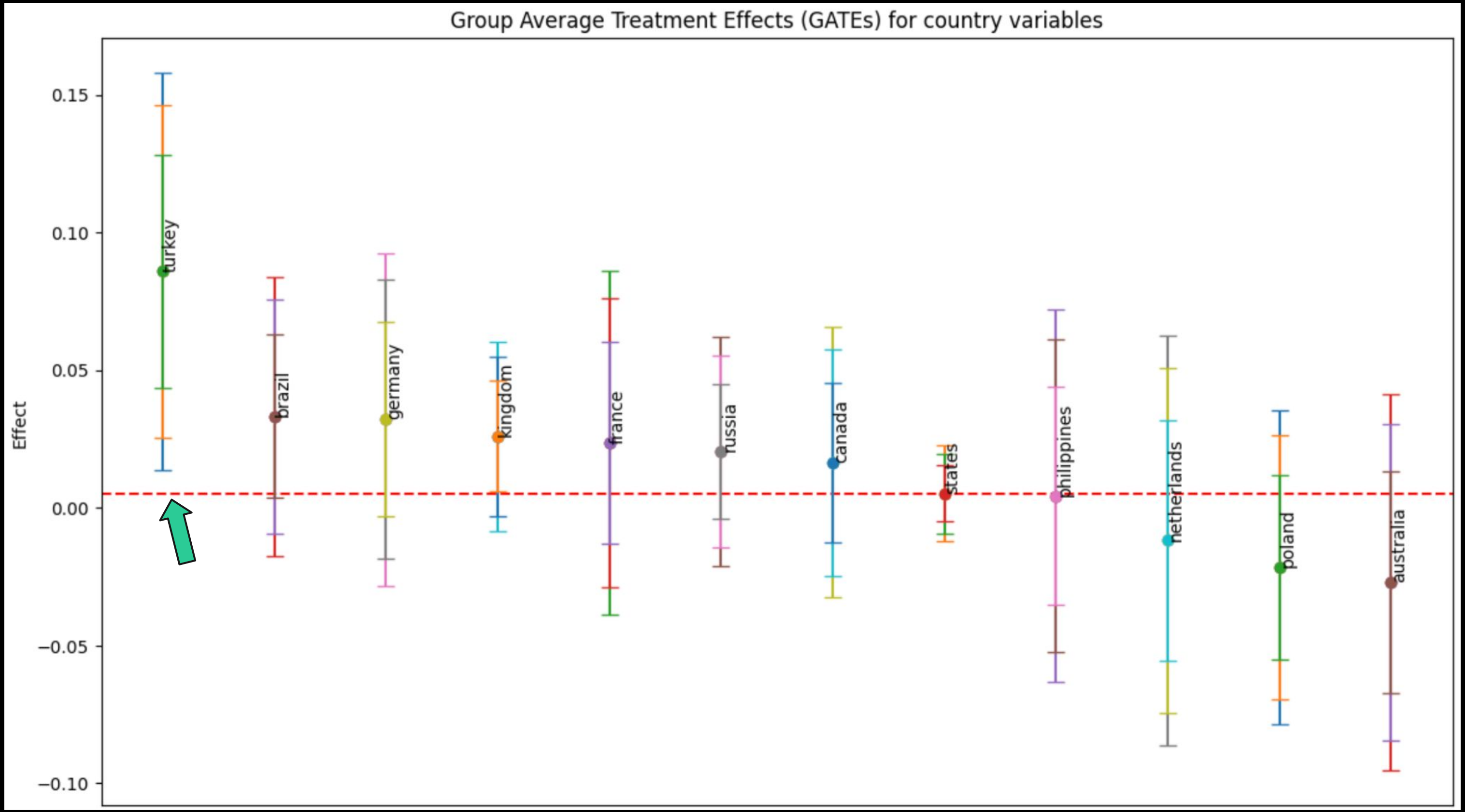
- 10 year olds
- 2-3 year tenure
- 25-50 universes



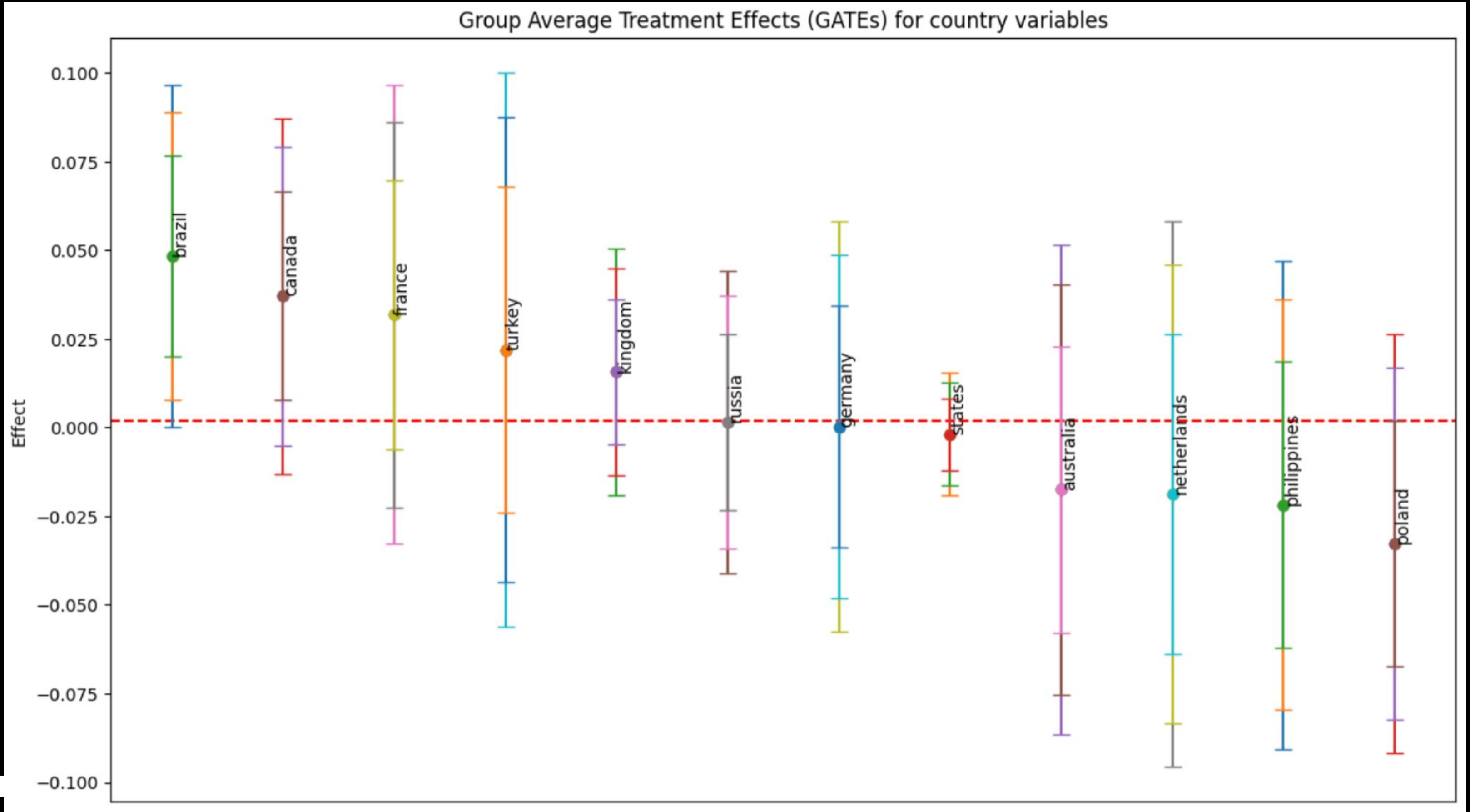
Method: DML-IRM

GPT: Countries

Pos stat sig impact in Turkey.

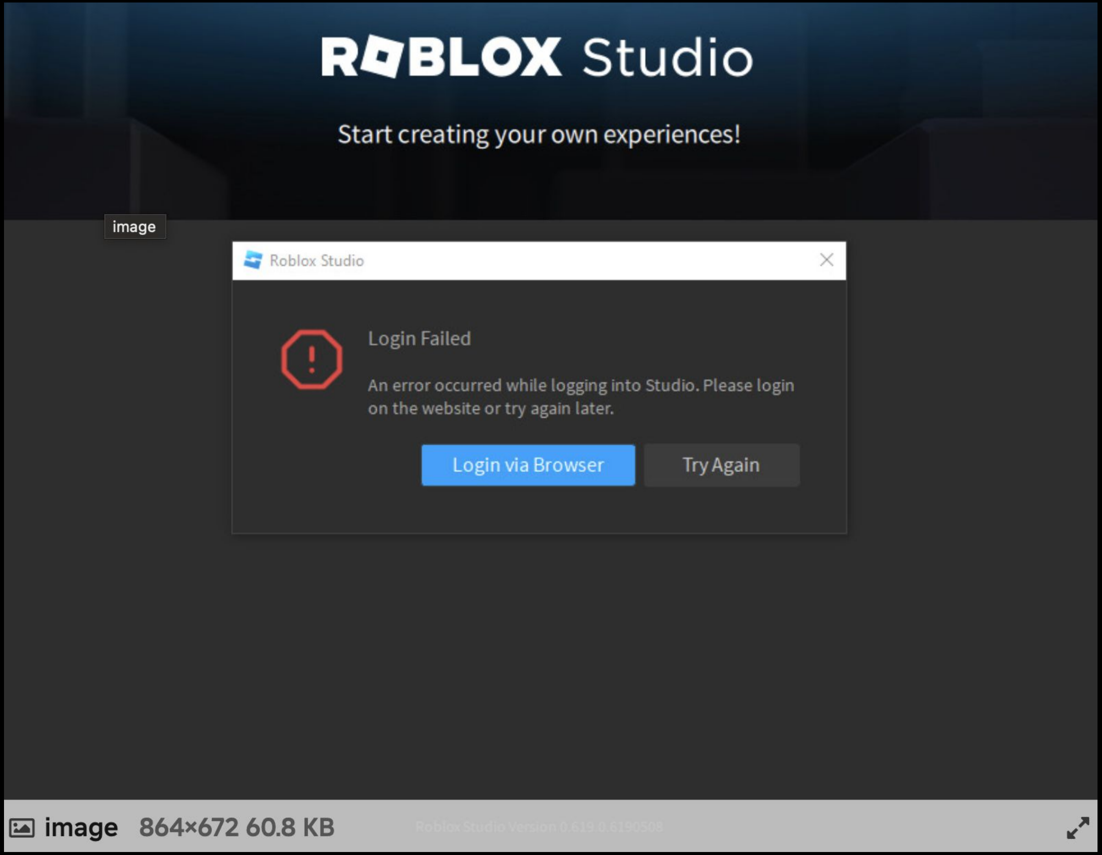


CodeLlama: Countries



Mac Bug Fixing

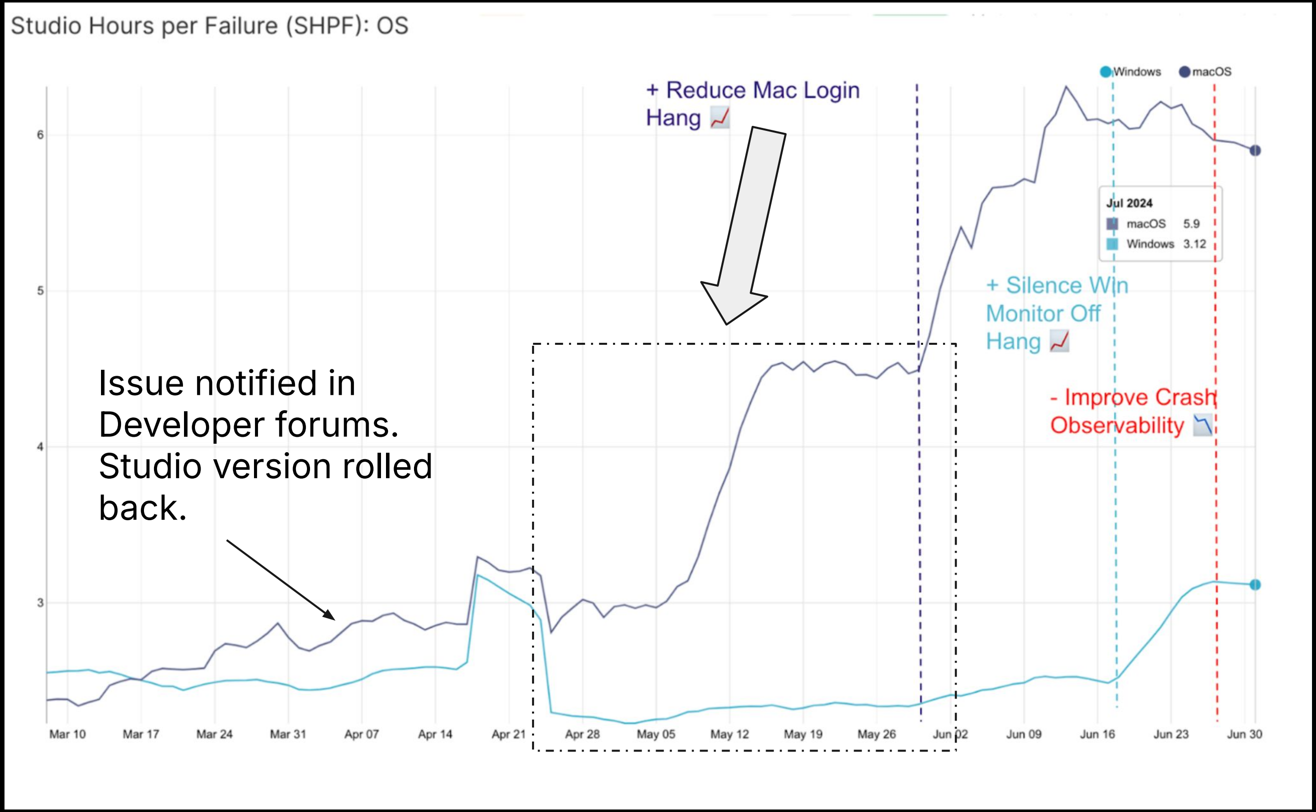
In April large number of Mac users began to face serious issues at login.



Mac Bug Fixing: Studio Hours Per Failure

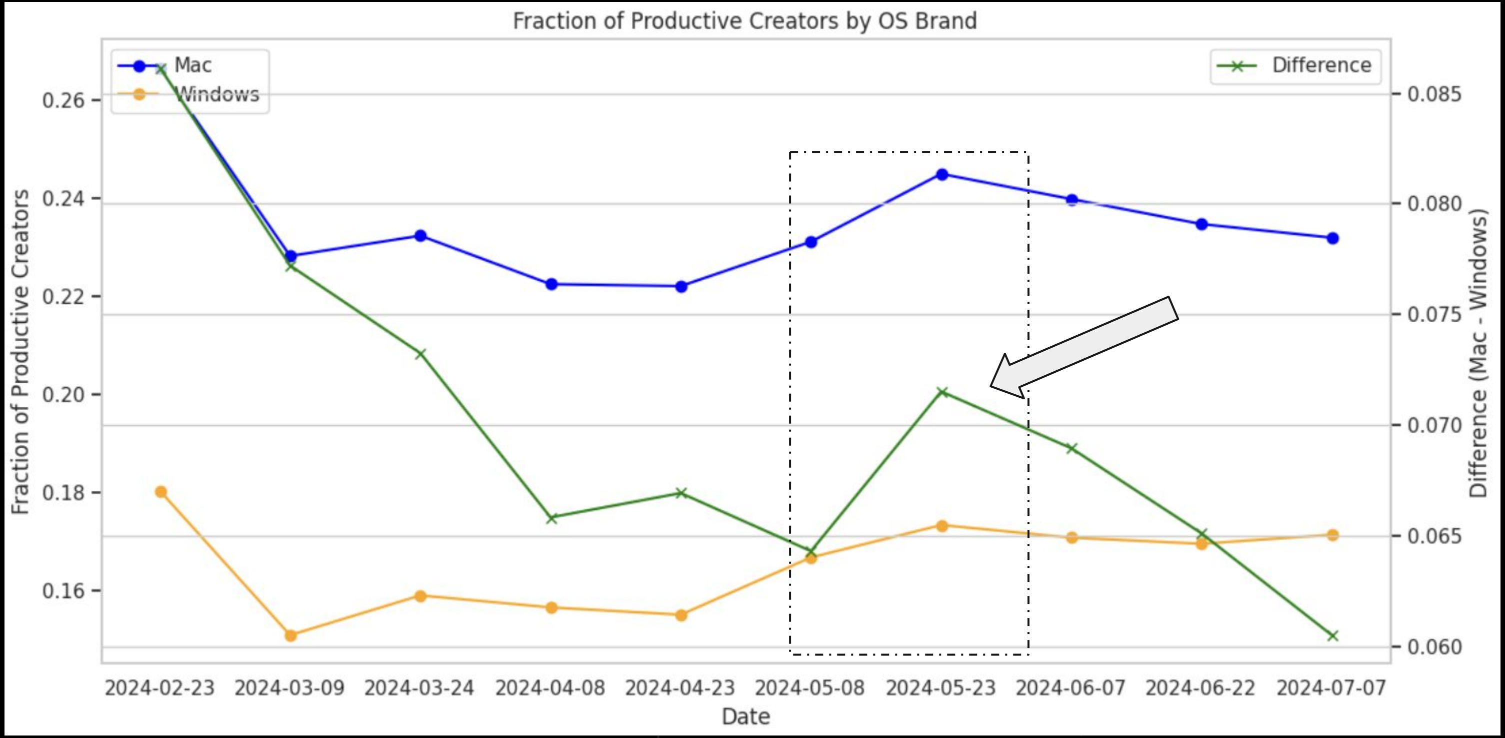
Focused bug fixing in mid-May led to approx 30% inc in the SHPF for Mac users.

Context: Engineers confirmed that no major update went out to Windows in May, and that this fix would majorly benefit Mac users, esp. who had not logged in for a long time.
 Dev forum link.
Git commit link.



Natural Experiment: Mac Bug Fixing

Declining Mac/window differences in productivity:

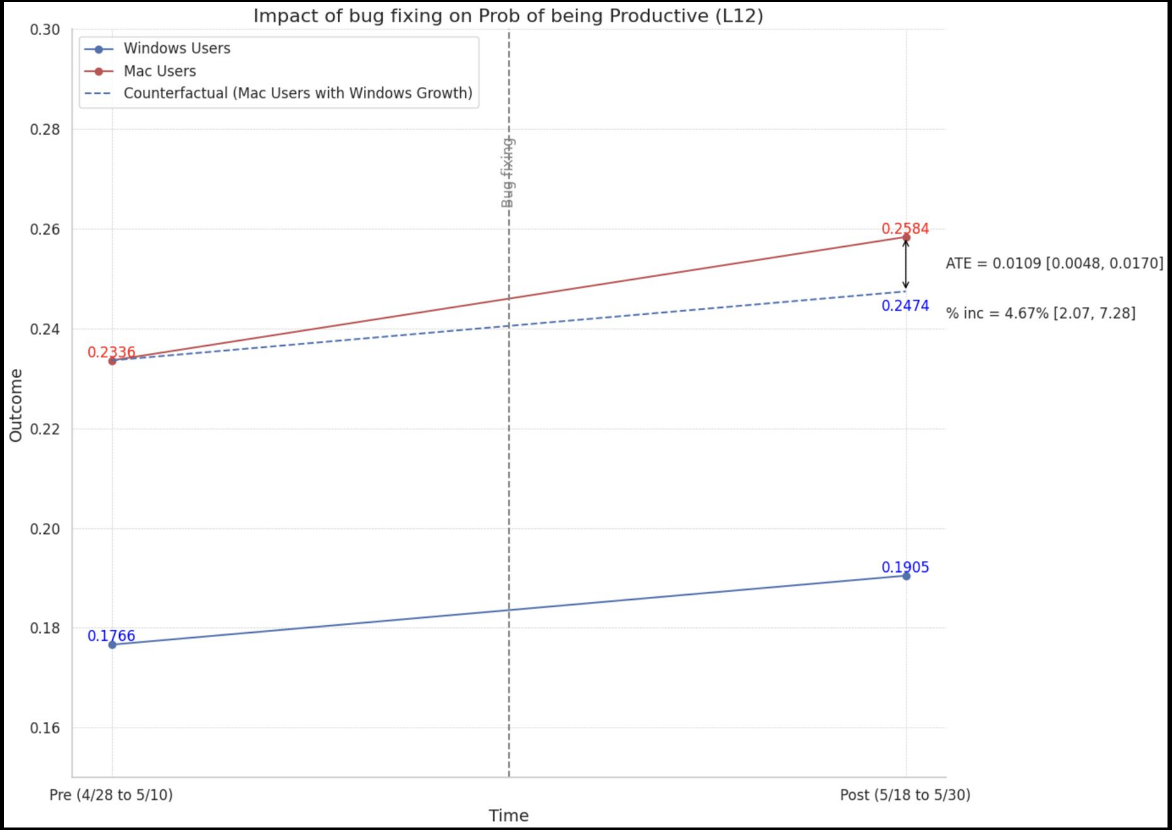


De15 day windows: 3 days to check eligibility, 12 days to check productivity.

Mac Bug Fixing: Instrumented Diff-in-Diff

A 5% stat sig increase in productivity in mac users, due to bug-fixing.

Thus a 10% increase in SHPF is associated with a 2% increase in Productivity in Mac users.



Suggested Enhancements

- Improvements:
  - Productivity not predictive of Bookings!
  - Thresholds could change over time or by genre.
- Linking productivity to creator milestones/movement through funnel.
- Measuring creator industriousness - code quality, efficiency, consistency, etc.
- Break up by actions (script, build, design, test, maintain, etc.)

Side Projects

- Observational studies of Gen AI Rollouts:
  - Material Gen, Texture Gen, Chat Assistant
- New techniques / designs:
  - Double ML + Diff in Diff
  - Genre Level Panel VAR
- Offline Evaluation for Reinforcement Learning Based Price Optimization
  - Bandits with Kalman Filter + Upper confidence bound