Universität St.Gallen

Country Roads, Take Me Home:
Evaluating Home Effect in the Tour de France

Eren Celik, Master Student in Economics at HSG
Xiaoyi Guo, PhD Student in Economics and Research Associate at SBF-HSG
Angelo Mimmo, PhD Student in Econometrics and Research Associate at SBF-HSG

From insight to impact.

Contents

Introduction and Background

Introduction

Home advantage in team sports: when home teams consistently win more than half of their games under a balanced schedule of home and away matches in a competition.

Size of home advantage: **heterogeneous** across sports, depending on the sport type and whether there is team cooperation (R. Pollard and G. Pollard, 2005).

Previous literature on individual sports:

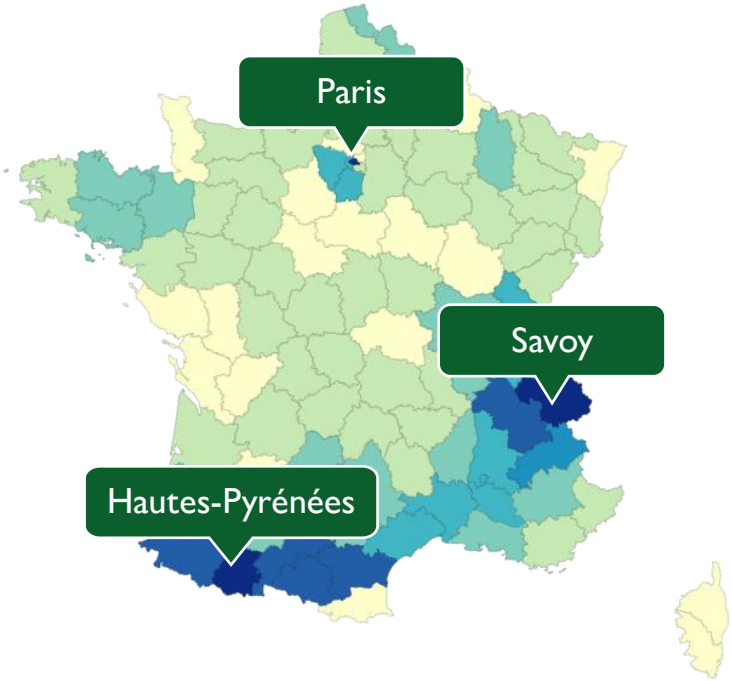
Reference	Sport	Approach	Results
Nevill, Holder, Bardsley, Calvert, and Jones (1997)	Golf and Tennis	Regressing tournament ranking on world ranking (both in logs) separately for home and away players	No significant home advantage neither in tennis nor in golf
Koning (2011)	Tennis	Logit model for the probability of winning a match	Significant home advantage exists for men, but not for women
Ramchandani and Wilson (2020)	Track and field athletics	Non-parametric statistical tests for difference in nations' medals and points under host and non-host conditions	Significant home advantage for indoor championships but not track championships

In the Tour de France, does riders' performance improve, decline, or remain unchanged when racing in their home country?

Background: Routes of Tour de France

Tour de France is one of the most important annual men multiple-stage cycling races (know as *Grand Tours*):

- Starting points: **not necessarily France**
  - Brussels (2019), Copenhagen (2022), Florence (2024)
- High chance of passing through the **mountainous regions** of France
  - Savoy (West Alps), Hautes-Pyrénées
- High chance of **final stage in Paris**



Background: Riders, Teams, and Stages

- The whole tour is split into **21 stages**:
  - Five types in three categories: **flat**, **hills**, **mountains**
  - In general, **1-2 time-trials** (ITT)
  - In general, **2 rest days**
  - Each stage has a winner, and the final winner is the one with the shortest cumulative time (i.e. the yellow jersey)
  - **UCI points for top 15 riders in each stage**
- **20-22 teams** competing:
  - **8 riders per team** participate in the tour
  - 18 UCI World Tour teams + "wild card" teams (UCI ProTeams)
  - All teams are **commercially sponsored**
  - Clear hierarchy in the team: **captain** and *domestiques*
- Both team and individual results are important:
  - **Short-term contracts** with high competition
  - Getting compensation to support the captain of the team
    - Money and/or support for the other races

Data and Variables Selection

Data

- Dataset: random sample of men's stages of Tour de France
  - Time period: **2020-2024**
  - Data granularity: **stage-rider**
  - Source: Pro Cycling Stats (PCS, www.procyclingstats.com)
  - Additional data collection: PCS and GeoPy (geocoding) APIs

- Data cleaning: we were able to keep **14.639 observations** out of 14.744 (only 105 lost)
  1. Total number of **stages for 2022** was 24 → imputed to 21
  2. Weight and height **missing** for some riders → retrieved them directly from PCS to compute BMI
  3. Two **duplicated columns**: *distance_km_tour* and *totaltdfdistance_tour* → dropped the latter
  4. Three riders **lacked publicly available weight and height** → removed
  5. One rider lacked the stage results for 2022 because retroactively **disqualified** → removed
  6. One rider did **not participate** in 2021 but we had (wrong) observations → removed

Variables

Type	Variable	Description	References
Outcome	*top15*	if a rider finishes within the top 15 in a given stage	
Treatment	*rider_home_country*	if the country of a given stage is equal to the rider's country	
Tour feature	*year_tour*	directly in the MCF and with dummies in the Logit model	
	distance_km_tour	total length of the Tour for a given year	
Rider's features	*age*	age in years at the starting date of the tour	Torgler (2007)
	bmi	body mass index, computed from height and weight	Torgler (2007); Phillips and Hopkins (2020)
	rider_points_previous_year	PCS points from the previous calendar year	Torgler (2007)
	rider_specialty	profile score specialty (i.e., "sprint", "climber" and "hills") with the largest number of PCS points	
	rider_specialty_tt	if TT points account for at least 50% of the maximum points across all the other categories	
	captain	first rider in the official start list of his team (number ends with 1)	Torgler (2007)
Team features	*team_points_previous_year*	sum of PCS points from the previous calendar year for all the riders in a team	Torgler (2007)
Stage features	*distance_stage*	distance of the specific stage in km	
	stage_type_RR	if it is a regular road race (RR, i.e. 1) or a time-trial (TT, i.e. 0)	
	profile_icon_stage	"Flat", "Hills, flat finish", "Hills, uphill finish", "Mountains, flat finish" and "Mountains, uphill finish"	
	perc_tour_completed	percentage of tour completed including that stage	
Combination of rider/team and stage features	*tt_match*	if the rider is a TT specialist and the stage is TT	
	specialty_match (_prev, _next)	if the profile score of the current, the previous and the following stages matches with the rider's specialty	
	area_knowledge	if the driving distance from the rider's birthplace to either the start or the end of the stage is ≤150 km	
	team_home_country	if the country of a given stage is equal to the country of the team	

Empirical Strategy

Empirical Strategy: Estimation Method

- Two estimators
  - Main estimation strategy: **Modified causal forest** (MCF, Lechner 2019; Lechner and Mareckova 2022)
  - Robustness check: **Logit** binary choice model (McCullagh, 1980)
- Advantage of MCF:
  - Identify **individual and group** average treatment effects
    - Home effect should not be homogeneous for all riders
    - Professional cycling has dispersed individualized properties
    - Professional cycling has strong heterogeneity in terms of riders, teams, and stages
  - No requirements of functional forms
- MCF estimates **IATE**, **GATE**, and **ATE**:

$$IATE(D;x) = \mathbb{E}[Y(D=1) - Y(D=0) \mid X = x]$$

Aggregate by group variation $Z = z$

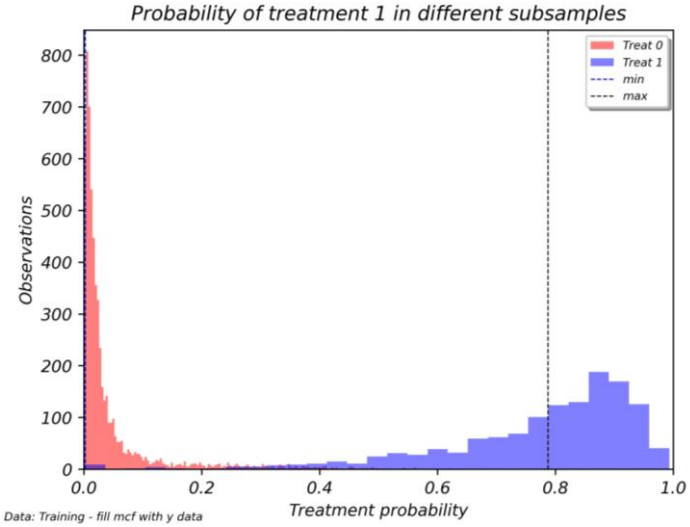
$$GATE(D;z) = \int IATE(D;\tilde{x},z) f_{\tilde{X}|Z=z}(\tilde{x}) \mathrm{d}\tilde{x}$$

Total aggregation

$$ATE(D) = \int IATE(D;x) f_X(x) \mathrm{d}x$$

Empirical Strategy: Identification

- **Conditional Independence Assumption**: well controlled variables ✓
  - Uncontrolled variables:
    - Schedule of the rest days → only 2 per tour, homogenous
    - Doping → very infrequent nowadays (Vandeweghe, 2022)



- **Common Support**: analyze data on common support sample ✓
  - 83.64% overlap (1,261 observations dropped)

- **Stable Unit Treatment Value Assumption**: ✓
  - No direct evidence from literature
  - Some spillovers are ruled out by design:
    - Psychological effects of HE not transferable
    - Team-level spillover exists, but we controlled for them
    - Mountainous stages cross also more remote area with less audience

- **Exogeneity of the Confounders** → next slide

Exogeneity of the Confounders: Team and Rider Selection

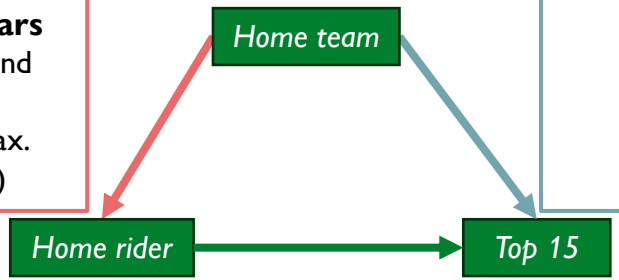
- **Exogeneity of the Confounders**: ✓
  - Potentially endogenous confounders removed:
    - Final Tour rank
    - Cumulative final time (and time gaps)
    - Cumulative bonus point obtained
  - *Team_home_country* problematic if rider team selection are correlated in some undesired way.

Treatment assumption

Domestic teams tend to choose domestic riders:

- In 2020, **1/3** of top teams' riders were **hired domestically** (Van Reeth 2022a)
- Teams have higher **bargaining power** than riders:
  - Average **contract** length of **1-2 years** (Phillips and Hopkins 2020; Larson and Maxcy 2013, 2016)
  - Very short professional **careers**, max. **2-5 years** (Larson and Maxcy 2016)

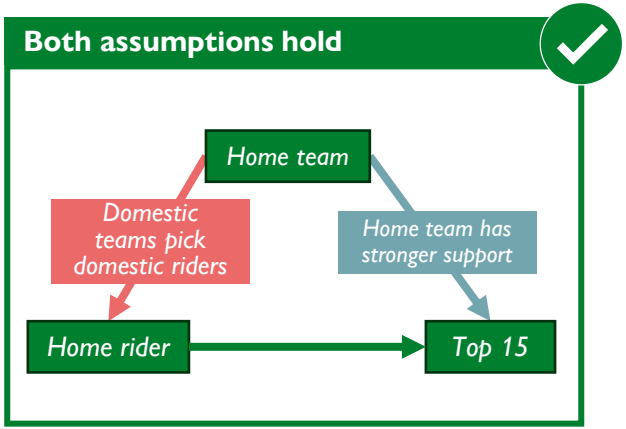
Directed Acyclic Graph (DAG)

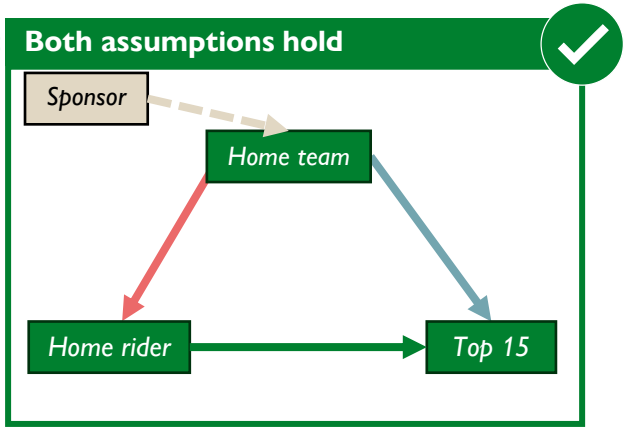


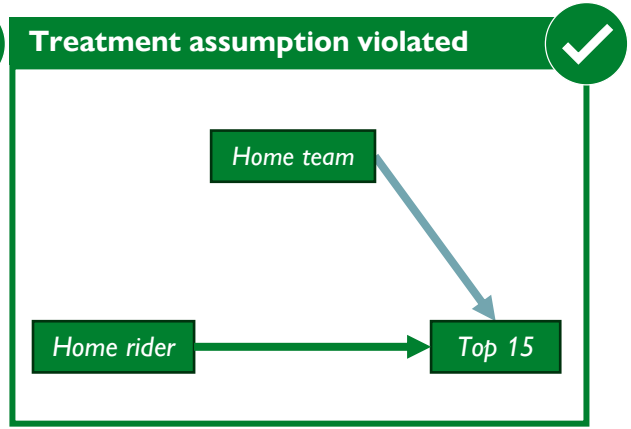
Performance assumption

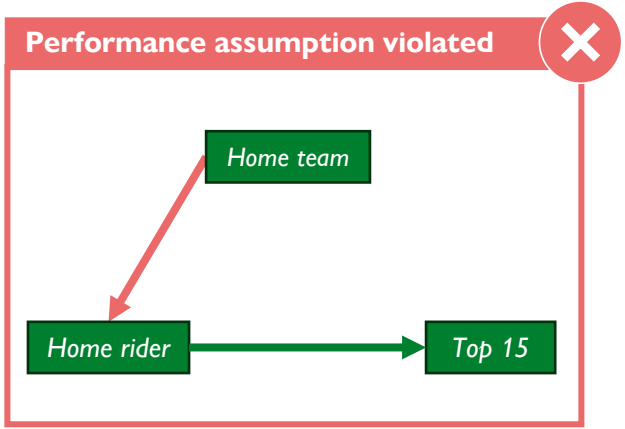
- Riders from home team feel stronger support from local fans
- Higher pressure by home teams on their riders
  - **2005 UCI ProTour reform** to avoid the problem of riders concentrating on competitions in their home countries or in the sponsor's home country (Rebeggiani 2016)

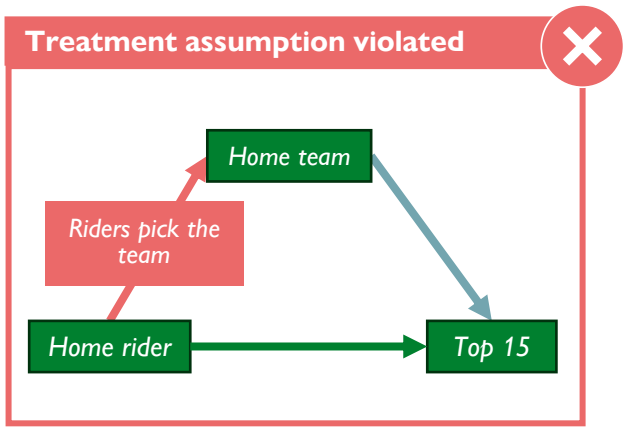
Exogeneity of the Confounders: Team and Rider Selection











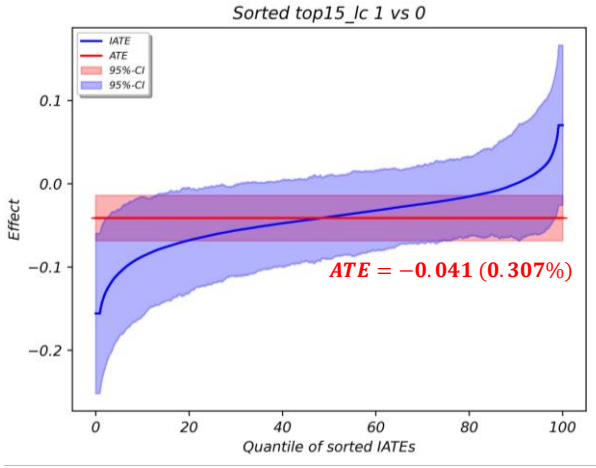
Results

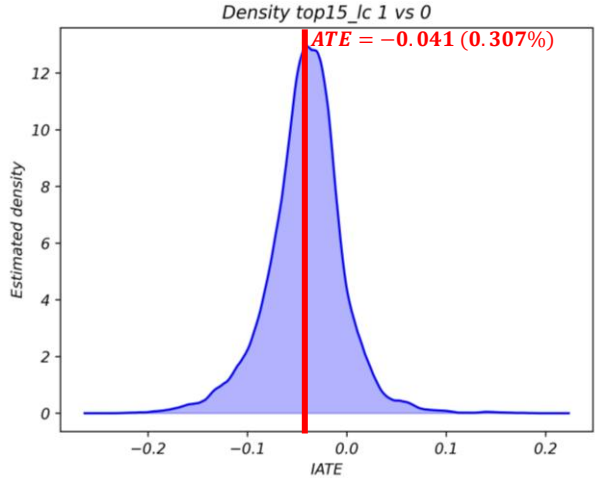
Balance Check: Common Support Sample

Variable	Full Sample		Common Support Sample	
	Untreated	Treated	Untreated	Treated
Area knowledge (%)	0.6	9.3	0.6	5.7
BMI (kg/m^2)	21.2	20.6	21.2	20.7
Captain (%)	6.8	8.8	6.9	4.8
Stage distance (km)	159.7	157.9	159.2	157.1
Tour completion (%)	50.7	53.2	50.5	52.2
Stage profile score	2.9	2.9	2.9	2.8
Stage rank	80.4	80.2	80.7	81.4
Rider's points previous year	479.9	392.3	481.0	438.6
Rider's specialty score	2.2	2.2	2.2	2.2
Rider's specialty TT (%)	21.5	10.6	20.3	12.6
Specialty match current stage (%)	35.7	34.3	35.7	35.6
Specialty match next stage (%)	33.6	33.4	33.6	32.0
Specialty match previous stage (%)	34.5	34.3	34.4	39.2
Stage type RR (%)	90.0	90.2	89.7	88.9
Team home country (%)	9.7	80.0	10.1	43.9
Team points previous year	3196.9	2646.2	3189.5	3218.2
Top 15 (%)	9.9	7.4	10.0	7.0
TT match current stage (%)	2.3	1.4	2.3	1.3

- The treated group has higher area knowledge (even with common support sample)
- We probably truncated the second-tier local teams/riders (the "wild card" UCI ProTeams invited)

MCF Results: Overview

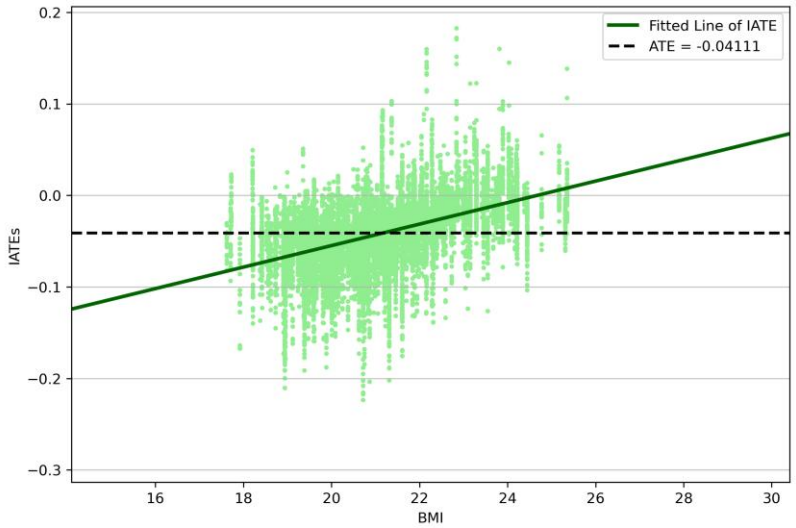




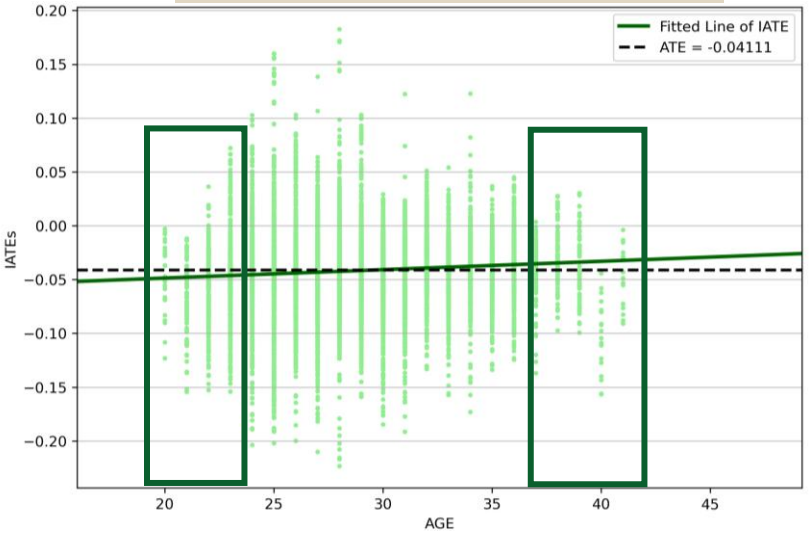
- Statistically **significant negative ATE** (-4.1%): **being a home rider reduces the probability of ranking into the Top 15**
- IATEs evenly distributed but we might have **heterogeneities** across individualized samples:
  - 9.98% of IATEs are positive
  - 37.65% of IATEs are statistically different from 0 at 5% significance level

    → *This motivates us to investigate IATEs and GATEs*
- Correspond to Böheim et al. (2019), Harb-Wu and Krumer (2019), Scoppa (2021), and Endo et al. (2023)

MCF Results: Individualized ATEs



Average drop out year: 2



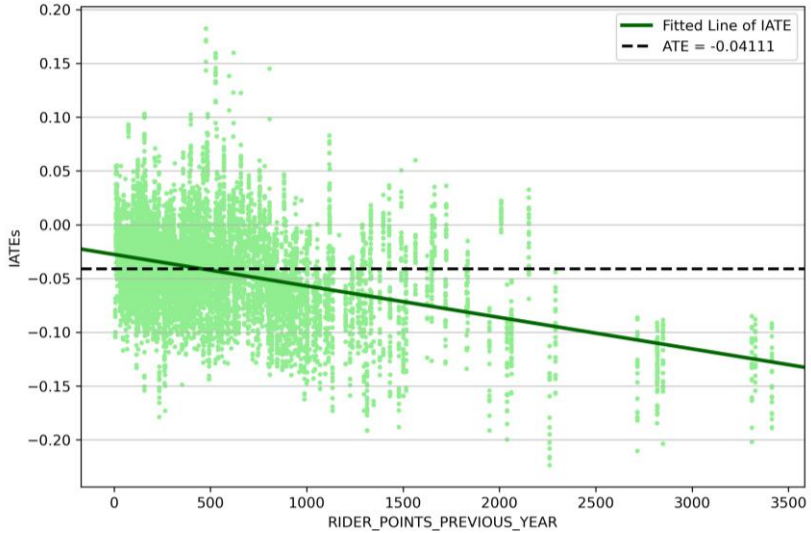
BMI

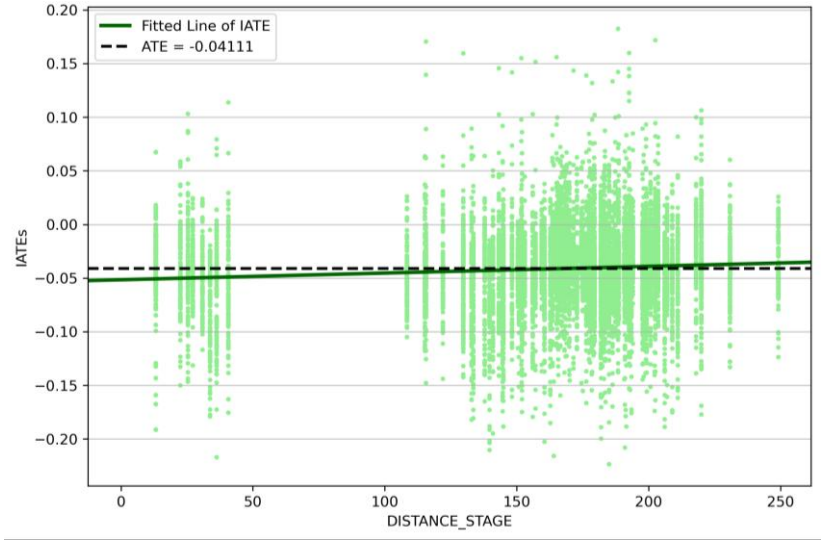
- Not conclusive
- No heterogeneity across individuals

Age

- No clear trend
- Effect variances for very young and older riders

MCF Results: Individualized ATEs





Rider's point last year

- Negative trend: more pronunced effect for higher points
- Pyschologically driven? Pressure?

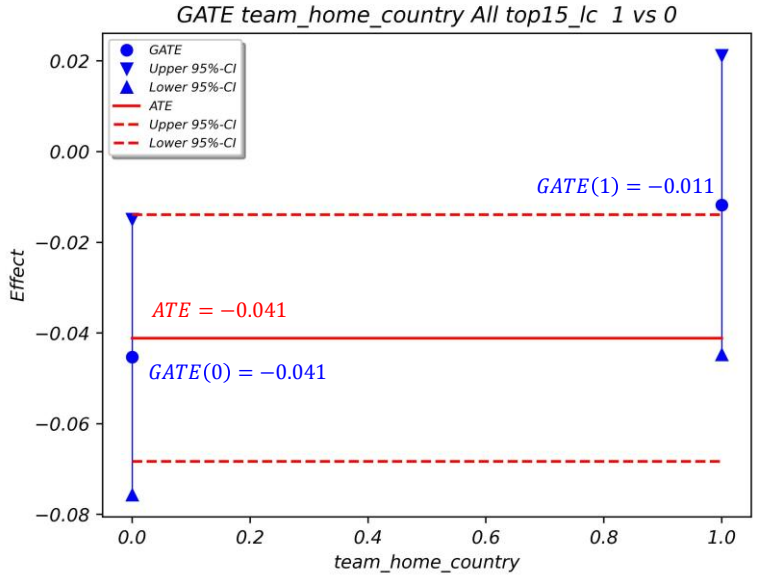
Distance of the stage

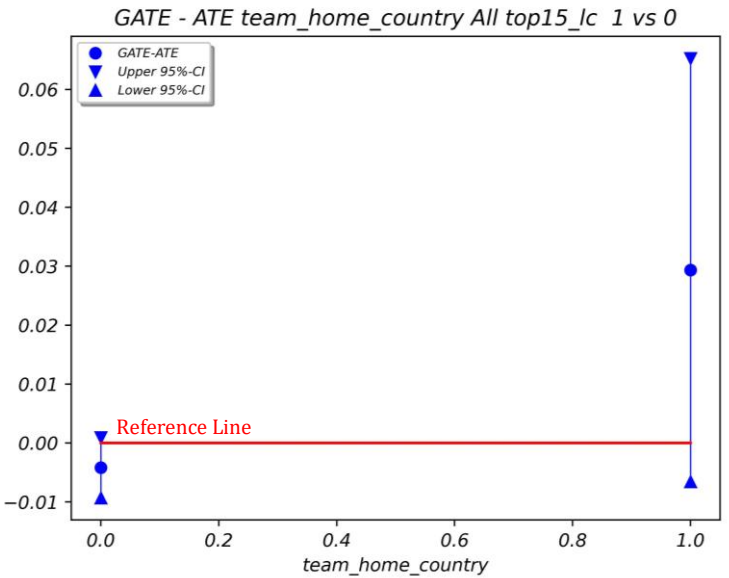
- Two clusters observed (TT and Regular)
- TT are slightly more affected (higher audience numbers)

MCF Results: Group Average Treatment Effect (Team Home Effect)

- We did not identify effect for home team status' heterogeneity to home effect.
- No difference to ATE.

Home Team	GATE	P-value	GATE-ATE	P-value
0	**-0.041**	**0.35%**	0.002	10.92%
1	-0.001	48.45%	0.018	10.95%

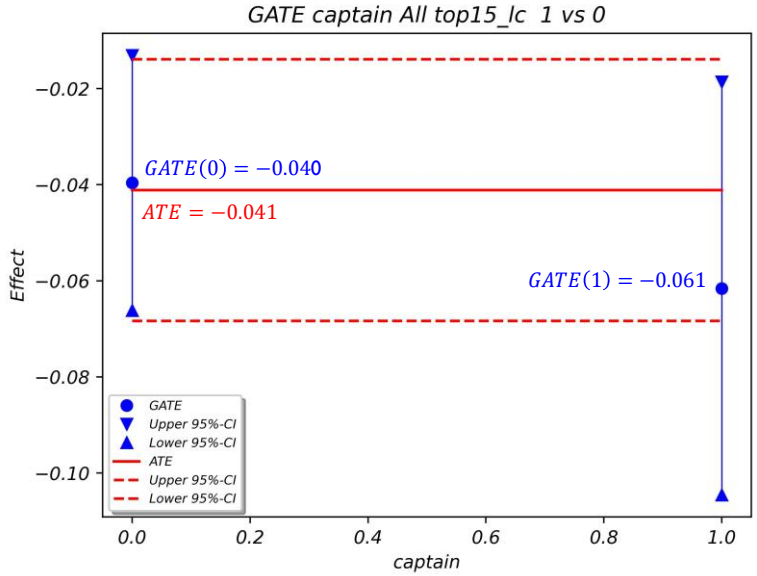


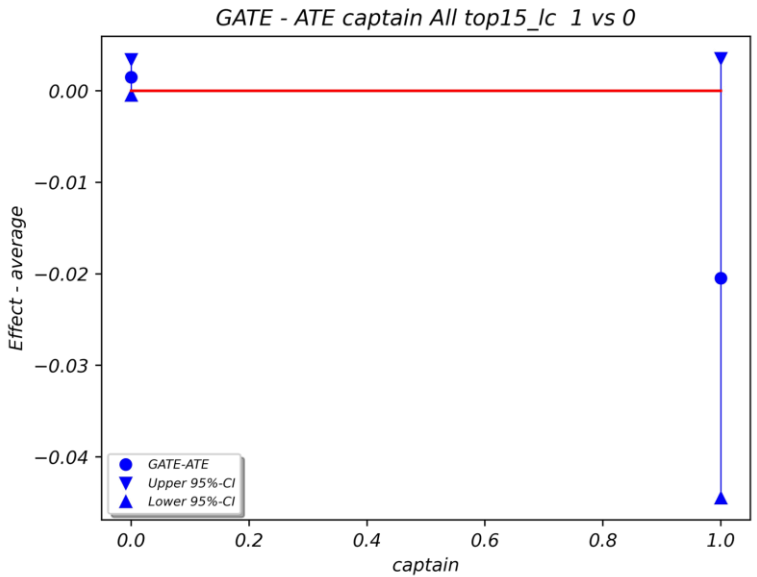


MCF Results: Group Average Treatment Effect (Captain)

- Being captain does not necessarily help rider overcome home effects better.

Captain	GATE	P-value	GATE-ATE	P-value
0	**-0.040**	**0.34%**	0.001	13.71%
1	**-0.061**	**0.50%**	-0.020	9.49%

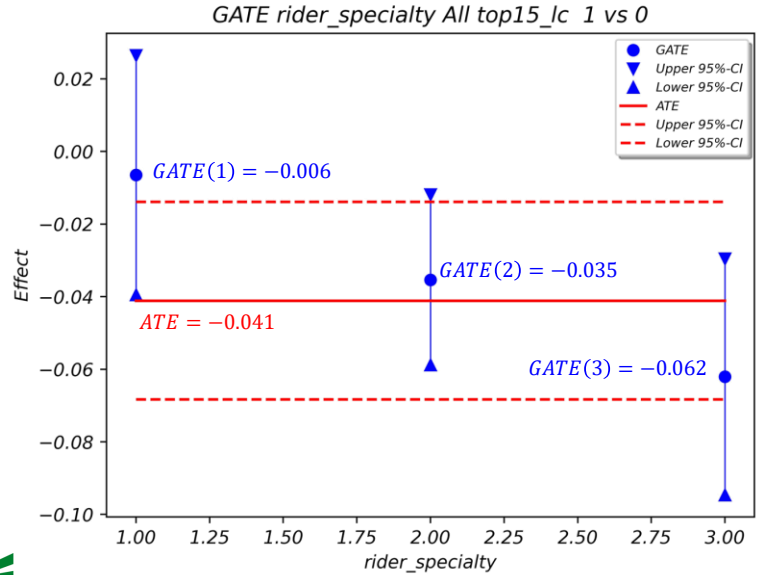


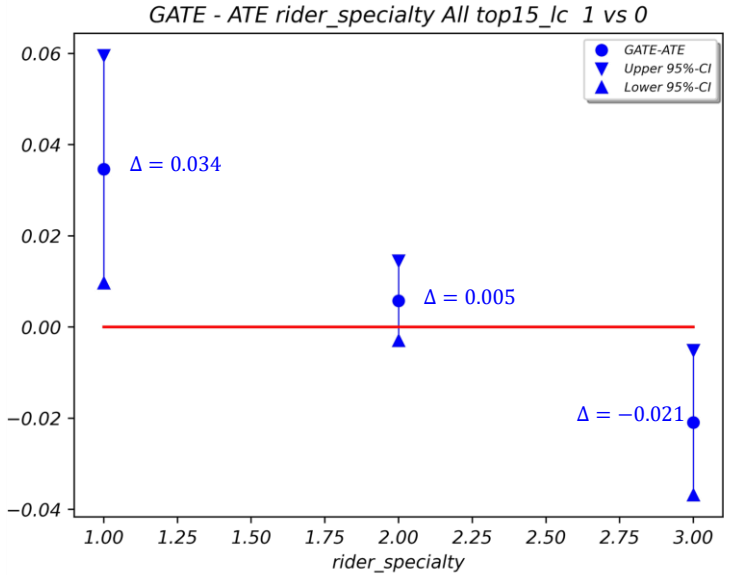


MCF Results: Group Average Treatment Effect (Rider's Specialty)

- Heterogeneities across specialties
- Sprinters and mountains experience less/more home effects
- Sprinters even have no home effect after all

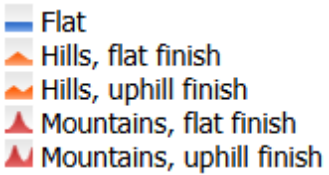
Specialty	GATE	P-value	GATE-ATE	P-value
1 = Sprinter	-0.006	69.95%	**0.034**	**0.66%**
2 = Hill	**-0.035**	**0.32%**	0.005	19.53%
3 = Mountain	**-0.062**	**0.02%**	**-0.021**	**0.97%**

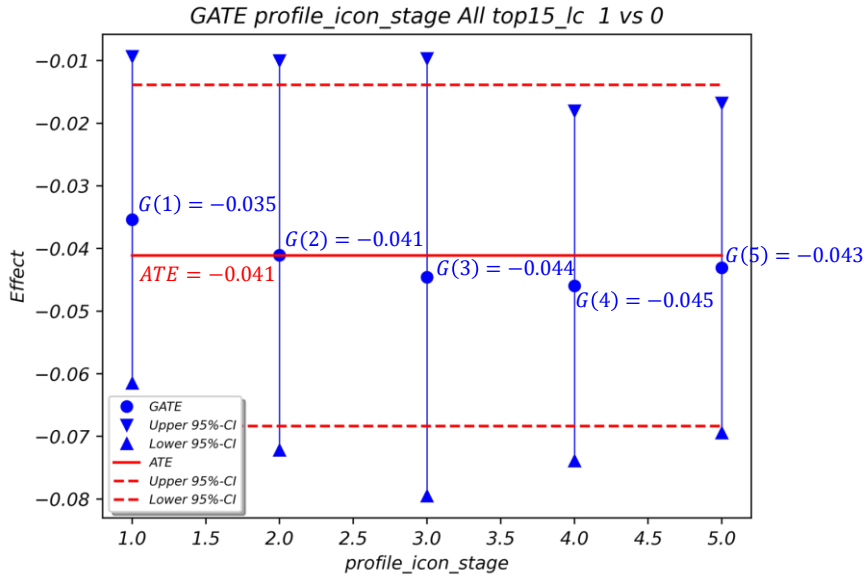


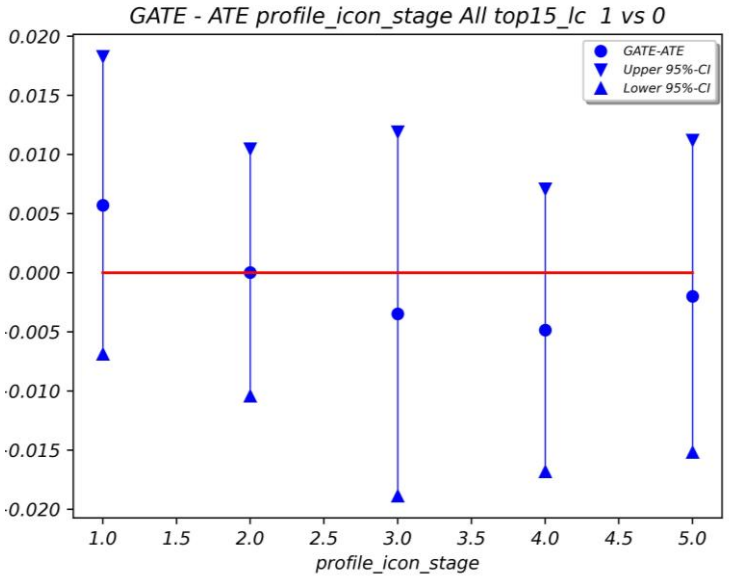


MCF Results: Group Average Treatment Effect (Parcours Profile)

- We identify GATEs for all different stage profiles (all statistically significant at 1% - 0.1%)
- But there is no difference between them and ATE
- The profile of a stage does not produce heterogeneity







Robustness Check: Logistic Regression

Variable	Marginal Effect (%)	*p*-value
year_tour_2021	3.09*** (0.66)	0.000
year_tour_2022	1.88** (0.60)	0.002
year_tour_2023	0.57 (0.60)	0.346
year_tour_2024	0.62 (0.62)	0.316
rider_points_previous_year	4.82*** (0.23)	0.000
bmi	0.17 (0.12)	0.164
captain	3.74*** (0.58)	0.000
age	–0.34*** (0.05)	0.000
team_points_previous_year	–0.33 (0.42)	0.438
distance_stage	0.21 (0.33)	0.520
profile_icon_stage	–0.44*** (0.13)	0.001
perc_tour_completed	0.02*** (0.01)	0.001
rider_home_country	–1.19 (0.84)	0.154
team_home_country	–1.60* (0.65)	0.014
tt_match	8.60*** (1.14)	0.000
specialty_match	7.99*** (0.39)	0.000
specialty_match_prev	–1.08** (0.39)	0.006
specialty_match_next	–0.69 (0.39)	0.079
area_knowledge	–2.29 (2.31)	0.321

- Captain: **positive** correlation
- Age: **slight negative** correlation

- This corresponds with our MCF results

- Rider home country: negative coefficient, but not statistically significant
- **Specification problem:** a rider finishing at 16th and one finishing at 150th are treated same
- The effort used from ranking 150-149 is *different* from that of 16-15

Conclusions

Conclusions

Home effect

Main result: negative ATE (**-4.1%**)

Interpretation: social and public **pressure**

Heterogenity

Effect strengthened for riders with **higher points** (i.e. better results, public exposure) in the past year

Effect eased for **sprinters**

Potential limitations

No direct measure or proxy for the pressure level

Treatment strongly concentrated on French riders

Thank you for the attention.

Eren Celik
Master Student in Economics
University of St. Gallen
eren.celik@students.unisg.ch

Xiaoyi Guo
PhD Student in Economics
University of St. Gallen
xiaoyi.guo@unisg.ch

Angelo Mimmo
PhD Student in Econometrics
University of St. Gallen
angelo.mimmo@unisg.ch

University of St.Gallen

Akkreditierungen

EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Bibliography

Böheim, René, Dominik Grübl, and Mario Lackner. 2019. "Choking under Pressure – Evidence of the Causal Effect of Audience Size on Performance." *Journal of Economic Behavior & Organization* 168 (December):76–93.

Brandenburger, Adam M., and Barry J. Nalebuff. 2011. *Co-Opetition*. Crown.

Breteau, Par Pierre. 2024. "Tour de France de 1903 à 2025 : quels sont les départements gagnants et perdants du tracé ?," November 3, 2024.

Correia-Oliveira, Carlos Rafaell, and Victor Amorim Anrade-Souza. 2022. "Home Advantage in Soccer after the Break Due to COVID-19 Pandemic: Does Crowd Support Matter?" *International Journal of Sport and Exercise Psychology* 20 (4): 1245–56.

Courneya, Kerry S., and Albert V. Carron. 1992. "The Home Advantage In Sport Competitions: A Literature Review," March. Earnheardt, Adam, Paul Haridakis, and Barbara Hugenberg. 2011. *Sports Fans, Identity, and Socialization: Exploring the Fandemonium*. Lexington Books.

Endo, Takuya, Hiroshi Sekiya, and Chiaki Raima. 2023. "Psychological Pressure on Athletes during Matches and Practices." *Asian Journal of Sport and Exercise Psychology*, Relationships between Physical Activity, Psychological and Cognitive Functioning: Evidence and Health Implications, 3 (3): 161–70. Fischer, Kai, and Justus Haucap. 2021. "Does Crowd Support Drive the Home Advantage in Professional Football? Evidence from German Ghost Games during the COVID-19 Pandemic." *Journal of Sports Economics* 22 (8): 982–1008.

Harb-Wu, Ken, and Alex Krumer. 2019. "Choking under Pressure in Front of a Supportive Audience: Evidence from Professional Biathlon." *Journal of Economic Behavior & Organization* 166 (October):246–62. h

Jones, Marshall B. 2013. "The Home Advantage in Individual Sports: An Augmented Review." *Psychology of Sport and Exercise* 14 (3): 397–404.

Koenker, Roger. 2005. *Quantile Regression*. Econometric Society Monographs. Cambridge: Cambridge University Press.

Koning, Ruud H. 2011. "Home Advantage in Professional Tennis." *Journal of Sports Sciences* 29 (1): 19–27.

Kroshus, Emily, Bernice Garnett, Matt Hawrilenko, Christine M. Baugh, and Jerel P. Calzo. 2015. "Concussion Under-Reporting and Pressure from Coaches, Teammates, Fans, and Parents." *Social Science & Medicine* 134 (June):66–75.

Larson, Daniel Joseph, and Joel G. Maxcy. 2016. "Human Capital Development in Professional Cycling." In *The Economics of Professional Road Cycling*,

Larson, Daniel, and Joel Maxcy. 2013. "Uncertainty of Outcome and Radio Policy in Professional Road Cycling." SSRN Scholarly Paper. Rochester

Lechner, Michael. 2001. "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption." In *Econometric Evaluation of Labour Market Policies* 43–58. 2019. "Modified Causal Forests for Estimating Heterogeneous Causal Effects."

Lechner, Michael, and Jana Mareckova. 2022. "Modified Causal Forest." arXiv.

Leota, Josh, Hoffman ,Daniel, Mascaro ,Luis, Czeisler ,Mark E., Nash ,Kyle, Drummond ,Sean P.A., Anderson ,Clare, Rajaratnam ,Shantha M.W., and Elise R. and Facer-Childs. 2022. "Home Is Where the Hustle Is: The Influence of Crowds on Effort and Home Advantage in the National Basketball Association." *Journal of Sports Sciences* 40 (20): 2343–52.

Lucía, Alejandro, Jesús Hoyos, and José L. Chicharro. 2001. "Physiology of Professional Road Cycling." *Sports Medicine* 31 (5): 325–37.

MacPherson, Ellen, and Gretchen Kerr. 2021. "Sport Fans Responses on Social Media to Professional Athletes Norm Violations." *International Journal of Sport and Exercise Psychology* 19 (1): 102–19.

Matthes, Julian, and David Piazolo. 2024. "Don't Put All Your Legs in One Basket: Theory and Evidence on Coopetition in Road Cycling." *European Economic Review*

McCullagh, Peter. 1980. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 42 (2): 109–42.

Menaspa, Paolo, Chris Abbiss, and David Martin. 2013. "Performance Analysis of a World-Class Sprinter During Cycling Grand Tours." *International Journal of Sports Physiology and Performance* 8 (3): 336–40

Mignot, Jean-François. 2022. "Strategic Behavior in Road Cycling Competitions." In *The Economics of Professional Road Cycling*, 19:227–51. Sports Economics, Management and Policy. Springer International Publishing.

Nevill, Alan M., Roger L. Holder, Andrew Bardsley, Helen Calvert, and Stephen Jones. 1997. "Identifying Home Advantage in International Tennis and Golf Tournaments." *Journal of Sports Sciences*, January.

Phillips, Kathryn E., and William G. Hopkins. 2020. "Determinants of Cycling Performance: A Review of the Dimensions and Features Regulating Performance in Elite Cycling Competitions." *Sports Medicine - Open* 6 (1): 23..

Pollard, R, and G Pollard. 2005. "Long-Term Trends in Home Advantage in Professional Team Sports in North America and England (1876 – 2003)." *Journal of Sports Sciences* 23 (4): 337–50.

Ramchandani, Girish, and Darryl Wilson. 2020. "Does a Home Advantage Effect Exist in International Track and Field Athletics?" *Managing Sport and Leisure* 25 (5)

Rebeggiani, Luca. 2016. "The Organizational Structure of Professional Road Cycling." In *The Economics of Professional Road Cycling*, edited by Daam Van Reeth and Daniel Joseph Larson, 33–54. Cham: Springer International Publishing.

Rubin, Donald. 1972. "Estimating Causal Effects of Treatments in Experimental and Observational Studies." *ETS Research Bulletin Series* 1972 (2): i–31.

Scoppa, Vincenzo. 2021. "Social Pressure in the Stadiums: Do Agents Change Behavior without Crowd Support?" *Journal of Economic Psychology* 82 (January):102344.

Torgler, Benno. 2007. "'La Grande Boucle': Determinants of Success at the Tour de France." *Journal of Sports Economics* 8 (3): 317–31.

Van Reeth, Daam. 2022a. "Globalization in Professional Road Cycling." In *The Economics of Professional Road Cycling*, edited by Daam Van Reeth, 337–67.

. 2022b. "The Finances of Professional Cycling Teams." In *The Economics of Professional Road Cycling*, edited by Daam Van Reeth, 35–67.

Van Reeth, Daam, and Daniel Joseph Larson, eds. 2016. *The Economics of Professional Road Cycling*. Vol. 11. Sports Economics, Management and Policy.

Vandeweghe, Hans. 2022. "Doping in Cycling." In *The Economics of Professional Road Cycling*, edited by Daam Van Reeth, 283–308.

Wikipedia. 2025. "List of Doping Cases in Cycling." In *Wikipedia*