DOI: 10.14754/CEU.2024.07

Essays in Applied Economics

by
Márton Fleck

Submitted to
Central European University
Department of Economics and Business

In partial fulfillment of the requirements for the degree
Doctor of Philosophy in Economics
Supervisor: László Mátyás

Budapest - Vienna,

2024

DOI: 10.14754/CEU.2024.07

Author: Márton Fleck
Title: Essays in Applied Economics
Degree: Ph.D.
Dated: June 2024

Hereby I testify that this thesis contains no material accepted for any other degree in any other institution and that it contains no material previously written and/or published by another person except where appropriate acknowledgement is made.

Signature of the author:

DOI: 10.14754/CEU.2024.07

Co-author contribution

Chapter 1: Favoritism under Multiple Sources of Social Pressure

Joint work with Gábor Békés and Endre Borza

The original idea of the chapter came from me. The focus of the chapter and the empirical strategy was developed in close cooperation with Gábor and Endre. The data was collected by Endre. I was responsible for the analysis coding. We wrote the text of the chapter together.

Chapter 2: The Xerox Effect: Communication Technology and Political Actions in Autocracy

Joint work with Arieda Muço

The original idea of the chapter came from Arieda. The data collection was done in close cooperation with Arieda. I was responsible for most of the data cleaning and the analysis coding. The first version of the manuscript was written by me, which was then substantially improved and modified by Arieda.

DOI: 10.14754/CEU.2024.07

Abstracts

Chapter 1: Favoritism under Multiple Sources of Social Pressure

Joint work with Gábor Békés and Endre Borza (Forthcoming, *Economic Inquiry*)

When social pressure leads to favoritism, policies might aim to reduce the bias by affecting its source. This chapter shows that multiple sources may be present and decomposing them is important. We build a novel and granular dataset on European football games and revisit the view that supporting crowds make referees help the host team. We find this bias to remain unchanged even in stadiums closed due to Covid-19. Instead, influential host organizations emerge as the source of social pressure. This has an adverse effect on maintaining the ranking of influential teams and hindering the progress of smaller teams.

Chapter 2: The Xerox Effect: Communication Technology and Political Actions in Autocracy

Joint work with Arieda Muço

In this chapter, we investigate the role that new communication technologies play in the fall of autocratic regimes. For this, we use a unique setting: the distribution of photocopy machines, the Xerox program, in communist Hungary. The photocopy machines were seen as a transformative technology, similar to the expansion and impact of the Gutenberg press. We use newly digitized data on machine allocation between 1985 and 1989 and show that areas with machines are more likely to participate in elections and support democratic values. Moreover, we show that adjacent areas are also affected. Our results suggest that new communication technologies help overthrow autocracies by promoting democratic values even when political competition is limited and traditional media is censored.

Chapter 3: The Causes and Consequences of International Student Migration and Mobility

The post-graduation stay-or-return decision of foreign students is crucial for the understanding of the welfare impacts of the internationalization of higher education. Students who return to their home country can contribute to the local economy with the skills acquired abroad. This can help global convergence. On the other hand, if students settle down in the host country permanently after graduation instead of returning, student mobility solely serves as a channel for 'brain drain' and thus contributes to growing inequality across countries. Using country- and university-level student flow data, I assess the factors that drive student mobility. The observed patterns are

DOI: 10.14754/CEU.2024.07

broadly consistent with the human capital model, which implies a relatively high return rate. The chapter also introduces a self-designed questionnaire. The responses to this questionnaire suggest a more nuanced picture, indicating that the majority of the students are not planning to return to their home countries in the future. Improving our understanding of this topic would be crucial for careful policy design.

DOI: 10.14754/CEU.2024.07

Summary

The thesis consists of three chapters, one single-authored and two co-authored. All chapters are based on original, novel datasets that help answer policy-relevant questions in various fields of applied economics. The individual chapters are summarized below.

Chapter 1: Favoritism under Multiple Sources of Social Pressure

Joint work with Gábor Békés and Endre Borza (Forthcoming, *Economic Inquiry*)

This chapter investigates potential mechanisms of social pressure leading to biased decisions. In particular, we are concerned with the context in which decision-makers such as judges and referees, despite being guided by strict rules, may regularly make biased and unfair decisions to preferentially treat one of the parties. We focus on cases when such favoritism is driven by social pressure – where influence is exerted by another person or group informally. This chapter extends a well-known exercise to decompose various sources of social pressure, looking at how football referees make biased decisions in setting the length of the stoppage time. According to a classic result, the pressure emanates from the supporting crowd, as referees internalize the preferences of the crowd in their decisions, by systematically favoring the home team.

With substantially finer data, we confirm the existence of a referee bias towards favoring the home team in European football, but identify a different source of this bias than previously assumed. Our exercise, covering granular data from the five top leagues over ten seasons (2011/12 to 2020/21), shows that the bias exists and referees support the home team by allowing the game to last longer if they are losing. The point estimate of this bias is 13 seconds or 5% of the extra time.

Our key contribution is understanding the source of social pressure pushing referees to make such decisions. Using empty stadiums owing to Covid-19 as source of external variation, we show that, unlike previously assumed, the bias is not due to crowds as the bias persists even in the absence of home fans. Instead, we show that the bias is attributed to the host team organisation, as the bias is exacerbated by the status of the hosting firm: influential teams benefit more than twice as much as the smaller ones (23 vs 11 seconds). Thus, we argue that influential host organizations emerge as the source of social pressure. This has an adverse effect on maintaining the ranking of influential teams and hindering the progress of smaller teams.

Chapter 2: The Xerox Effect: Communication Technology and Political Actions in Autocracy

Joint work with Arieda Muço

DOI: 10.14754/CEU.2024.07

This chapter investigates the role that new communication technologies play in the fall of autocratic regimes. For this, we use a unique setting: the distribution of photocopy machines (called the Xerox program under the umbrella of the Soros Foundation) in communist Hungary. The photocopy machines were seen as a transformative technology, similar to the expansion and impact of the Gutenberg press. We assess the extent to which this new technology helped to promote democratic values and contributed to the fall of the communist regime, resulting in a democratic regime change by 1989/90.

We use newly collected and digitized data on the allocation of Xerox machines to public institutions in Hungary between 1985 and 1989. Using various empirical strategies such as propensity score matching and staggered difference-in-differences, we show that the Xerox program had both short-term and long-term effects: in geographic areas equipped with machines, people were more likely to support democratic values in the 1989 referendum, participate in the 1990 general elections, and finally, less likely to support the 2016 anti-Soros referendum of the Hungarian government. We also show that adjacent areas are also affected. Our results suggest that new communication technologies help overthrow autocracies by promoting democratic values even when political competition is limited and traditional media is censored.

Chapter 3: The Causes and Consequences of International Student Migration and Mobility

Improving the international mobility of higher education students has been a major policy priority agenda of the European Union. The overall figures suggest that the project can be considered successful, as the number of degree-mobile students has been in a constant rise over the last twenty years. In the economic literature, generally two broad theories are used to model student mobility, and these come to different predictions as to the motives and consequences of student migration. According to the human capital model, student migration takes place to acquire skills that are not available at one's come country. In this case, the student chooses destination based on education quality, and is likely to return to home country after graduation (or settle down elsewhere). The migration model, on the other hand, considers student migration as a tool for settling down permanently in the destination country. This model predicts that future employment prospects and general conditions of the destination country are more important than university quality.

The aim of this chapter is to empirically assess the drivers and consequences of international student mobility. To give a comprehensive account, I investigate these at three levels. First, a country-level analysis sheds light on how macroeconomic and geographic factors of origin and destination countries affect student migration flows. Using a gravity model estimation framework built on dyadic country level panel data, I test the predictions of the two alternative economic models. The results suggest that the student flows are more consistent with the human capital model, which predicts a relatively higher return rate. As a second step, a university-level analysis

DOI: 10.14754/CEU.2024.07

is carried out based on the European Tertiary Education Register (ETER) project's comprehensive data collection of European Union universities between 2017 and 2020. These results shed light on how various university level factors such as tuition fees, student-to-faculty ratio, research quality, and university rankings affect the number of international students. These results broadly confirm the empirical validity of the human capital model as opposed to the migration model: students tend to focus more on university quality (proxied by ranking scores) than the destination countries' feature.

Finally, I carry out an individual level analysis. As part of this chapter, a web-based survey was designed to uncover international students' motives and post-graduation plans. Based on the responses collected from 600 international CEU students, I assess whether the results of the country- and university-level analyses are in line with the individual responses. Furthermore, this data allows to focus on an aspect that is completely missing from the existing datasets and literature: the likelihood of the students' returning to their home country after graduation, along with the factors that affect their return decisions. The responses to this questionnaire give a more nuanced picture than the country- and university-level analyses, as a slight majority of the respondents claim that they are not planning to return to their home country in the near future. This result has important major policy implications, as higher education mobility, when combined with low return rates, arguably contributes to brain drain and thus may inhibit the convergence of less developed countries.

DOI: 10.14754/CEU.2024.07

Acknowledgements

I would like to express my gratitude to my supervisor, László Mátyás for his support and guidance throughout my doctoral studies. This thesis would not have been possible without his help.

I would like to thank my co-authors, Arieda Muço, Gábor Békés, and Endre Borza for the inspiring conversations that led to our joint projects, and the time we spent on bringing these papers to life.

I am indebted to the examiners of my thesis, Mark Harris and Sergey Lychagin for the extensive and invaluable feedback they provided, as well as to Péter Benczúr and Julius Horváth for their work as members of the Thesis Committee.

I am grateful to Miklós Koren, Ádám Szeidl, and all my ex-colleagues at the CEU Microdata research group. The data skills I acquired during the years we worked together were essential for all chapters of this thesis.

Finally, I am most grateful to my family, especially my wife Juli and my children Emil and Liza for their love, emotional support, and patience.

Contents

List of Tables

List of Figures

DOI: *DOI: 10.14754/CEU.2024.07*

1 Chapter 1: Favoritism under Multiple Sources of Social Pressure

Joint work with Gábor Békés and Endre Borza

1.1 Introduction and background

In competitive situations between organizations, an agent (such as a judge or a referee) often has the opportunity to make a decision benefiting one party. Favoritism occurs when a decision-maker gives preferential treatment to one party at the expense of another, without justification from rightful determinants. Favoritism appears in various contexts, from promotion decisions in organizations (Prendergast and Topel, 1996) to the allocation of regional development funds (Hodler and Raschky, 2014).

Beyond explicit corruption (bribery), another common source of favoritism is social pressure: the exertion of influence by a person or a group. Social pressure includes conformity and persuasion. Conformity involves agents' desire to conform to expectations, seeking approval or adhering to a social image of themselves (Bursztyn and Jensen, 2017). Persuasion occurs when individuals or groups influence an agent's decision in their interest via persuasive behavior such as communication (DellaVigna and Gentzkow, 2010), looking for both cognitive and emotional responses from the receiver (DeMarzo et al., 2003; Schwartzstein and Sunderam, 2021).

When social pressure leads to favoritism, policy actions should raise awareness and aim to reduce bias by affecting either party (the sender or the receiver). This is straightforward when there is a single channel of social pressure. However, individuals may care about the perception of more than one reference group (Bursztyn and Jensen, 2017). For instance, students may care about both the perceptions of their peers and those of their prospective employers (Austen-Smith and Fryer, 2005). Regarding persuasion, consumers might be targeted by various parties simultaneously.

We contribute to the literature by showing the importance of distinguishing sources of social pressure. We revisit Garicano et al. (2005), who analyzed the decisions of association football (soccer) referees and found evidence of favoritism, specifically referees favoring the home team in crucial situations.[1] They argued that this bias is driven by social pressure from the home team supporters in the stadium. We argue that referees may be influenced by multiple sources of social pressure, such as the supporting crowd and the host team organization. Using exogenous variation in crowd size due to stadium closures during the Covid-19 pandemic, we find no effect from crowds on referee behavior. Instead, we show that the host team's organization influences referees. In particular, more successful teams have greater influence and benefit more from referees' favoritism

[1]There is a growing literature using sports to learn about behavior. For instance, the expectation of financial rewards was shown to lead to match rigging in Japanese Sumo wrestlers (Duggan and Levitt, 2002). Gauriot and Page (2019) also uses football data to discuss quality perception biased by luck, while Parsons et al. (2011) looks at discrimination in US baseball.

 DOI: *DOI: 10.14754/CEU.2024.07*

bias. Nevertheless, social pressure remains the driver of favoritism, as we find no evidence of corruption: the bias in favor of successful teams is uncorrelated with referee career paths.

To differentiate these sources, we need appropriate data and an identification strategy. We compile a comprehensive dataset on European football games covering match events, referee decisions, and referee career paths. Our high granularity dataset spans 10 seasons (2011/12 to 2020/21) of the 5 most prestigious men's football leagues: the English *Premier League*, the Spanish *La Liga*, the Italian *Serie A*, the German *Bundesliga 1*, and the French *Ligue 1* first divisions. We work with event-by-event data that records each action (such as passes, disciplinary actions, penalties, and injuries) along with a timestamp.

The data covers a long period before the Covid-19 pandemic as well as games played during the pandemic.

This paper confirms the existence of favoritism as shown by Garicano et al. (2005): referees use their discretionary power to systematically award more stoppage time to the home team when it stands to benefit.[2] Comparing games with a one-goal difference after regular time, we find that referees, on average, add 13 seconds more stoppage time when the home team is losing compared to when the home team is winning.[3]

Next, we explore the sources of social pressure and examine the role of supporting crowds. Garicano et al. (2005) argue that referees' home bias is driven by social pressure from home team supporters in large numbers in the stadium: referees conform and internalize the preferences of home team supporters. This mechanism is identified via variations in crowd size and its composition: larger attendance generates greater bias, while a higher attendance-to-capacity ratio—indicative of a higher share of visiting crowds—yields lower bias. These findings are extended by Dohmen (2008), who suggested that the distance from crowds—in stadiums with running tracks—moderated the influence of crowds.

We find, however, that the bias is not driven by crowds. The global Covid-19 pandemic in 2020 and 2021 led to sudden and unexpected stadium closures, and games were played in empty stadiums. Instead of relying on observational variation in the relative size of home supporters, we use stadium closures as a natural experiment. Using this exogenous variation in the home support by local crowds, we found that the bias remained unchanged even without supporters in the stadium.

The home team's organization is the second source we investigate. The organization hosts the referees, who might experience social pressure from them. Persuasion could occur with the host team having the chance to repeat a positive message, and hospitality may play an important role. On match day, referees are surrounded by members of the host organization and often spend the

[2]The discretionary power empowers referees to set stoppage time at the end of each half to compensate for time lost due to various events such as injuries. See Section A.1 of the Appendix for more details.

[3]This difference corresponds to 5% longer stoppage time, or an average of one additional point per season and team. This is somewhat smaller than the 20-second bias found by Dohmen (2008) in the German Bundesliga for the 1992/93 to 2003/04 period, and substantially smaller than the 110 seconds Garicano et al. (2005) found for the two Spanish seasons of 1994/95 and 1998/99.

whole day at the premises. Before, during, and after the match, the home team staff looks after the safety and well-being of the referee. Gifts from the home team are not uncommon either.[4]

At the same time, referees in close proximity to players and managers could be the strongest channel. Players, especially stars, will confront referees on decisions they do not like. Famous players and managers will exert star power. Both the persuasion by these individuals and the human desire to conform to greatness could play a role.[5] Furthermore, players, when ex-ante expected to win at home, could exert higher levels of pressure on the referees when losing or drawing as they themselves are feeling the pressure.[6]

Social pressure is likely to stem from both teams, but its relative size will vary with influence. In football, the size of influence is related to sporting success (such as ranking after a season). For referees, the desire to conform to successful teams and their star players will be stronger. These teams will also mobilize more resources (better facilities and personnel) for persuasion.[7] To detect social pressure related to the host team organization, we compare teams of different rankings: more successful teams are expected to enjoy a larger home team bias.

Looking at the difference in stoppage time when it benefits the home team compared to when it benefits the away team, we indeed find that referees add more than twice as much stoppage time at the end of the second half for the benefit of the most successful third of teams compared to the rest.[8] The favoritism bias is the largest when a top-ranked team plays a minnow: the lowest-ranked teams will enjoy zero home bias when playing against top teams.

In our exercise, we aimed to show an example where social pressure comes from different sources. We found that the traditionally assumed source, namely crowds, cannot explain favoritism. How can we reconcile our results with earlier ones? We suspect that earlier results on crowds may have been confounded by the characteristics of the host team organization that remain in play even behind closed doors: more successful teams have larger stadiums and are less likely to have running tracks (no top team in Germany has tracks, while 9.8% of the rest have tracks). A large and highly granular dataset and exogenous variation were necessary to distinguish these two sources of social pressure.

Importantly, social pressure may have an aggregate consequence in our setup. If the top-ranked teams get additional help from referees, social pressure will contribute to maintaining the ranking, making it harder for smaller teams to catch up.

Finally, we examine whether a more obvious explanation for favoritism could also act as a

[4]For anecdotal evidence, see www.theguardian.com/football/2015/mar/28/referees-football-match-day-routine-sport and www.as.com/en/2017/02/27/soccer/1488229755_283818.html.

[5]Conformity has a large literature in psychology, see Cialdini and Trost (1998), with conformity to stars, in particular, discussed in many contexts such as body image in Shorter et al. (2008).

[6]We thank an anonymous referee for this point.

[7]Sporting success and team wealth are strongly correlated in elite sports such as football, and there is low churning at the top.

[8]This is in line with a related finding: using an expert panel, Erikstad and Johansen (2020) showed that the top two teams in Norway are more likely to get a penalty awarded.

 DOI: *DOI: 10.14754/CEU.2024.07*

key possible confounder: corruption. While direct bribery is extremely unlikely,[9] referees may expect professional rewards in the future when helping the most successful teams.[10] To refute this alternative hypothesis, we examined referees' careers in the prestigious pan-European competitions where only the most successful teams play. If favoritism was driven by the expectation to work more in these leagues, referees with higher bias would be picked more often – in collusion with participating teams. We find no such evidence.

Several previous studies have looked at sports outcomes in games played behind closed doors, without supporting crowds. Bryson et al. (2021), Reade et al. (2022), Wolaver and Magee (2022), and Morita and Araki (2022) all found that in terms of disciplinary action, the home bias changed substantially during closed games. However, while such actions are easier to observe, identification is problematic, as these referee decisions are intertwined with the actions of the players. This makes it difficult to empirically disentangle referee and player behavior.[11] Importantly, Caselli et al. (2022) identified a differentiated effect of crowd presence on player behavior and performance. Specifically, African players performed better in closed games. This indicates that support or opposition of attending crowds can have diverse and specific impacts on individuals participating in the games.

Another related paper is Gong (2022), who assesses whether referees' home bias changed without crowd support in NBA games. He compares the probabilities of incorrect referee decisions made against home versus away players in the last two minutes of narrow-margin games. He finds that the empty arenas due to the Covid-19 pandemic have no differential effect on home bias. However, in this basketball example, no baseline home bias was identified even before closed games. This means that the confounders of the referees' bias, such as team influence, could not be investigated either.

This paper contributes to understanding that social pressure may be related to several actors simultaneously, with different policy outcomes. Consider possibly biased arbitration judgments in investor-state dispute settlements (Behn et al., 2018). There may be various social pressures: persuasion by and conformity to wealthy countries and crowd pressure via social media or protests. In the case of investigating bias in online reviews (Vollaard and van Ours, 2022), reviewers may be subject to social pressure when favoring big brands popular among large sections of consumers, as well as be targets of persuasion by companies sending gifts and offering marketing events. In both these scenarios, our results suggest that analysis must tackle different sources rather than assuming any. The paper is also related to the literature on multi-sender communication (Battaglini, 2002; Gentzkow and Kamenica, 2016), showing a case of strongly correlated signals by the senders.

As we are interested in broader social settings, we acknowledge that the world of football is

[9] Although not unheard of, see the Italian match-fixing case https://en.wikipedia.org/wiki/Calciopoli

[10] For instance, the most successful teams are able to pressure the UEFA to adapt its competition formats in their favor or avoid penalties. For some discussion, see Appendix Section A.1.5.

[11] See e.g. Carmichael and Thomas (2005) and Dawson and Dobson (2010), and for a review, Dohmen and Sauermann (2016).

particular. However, all games are televised, and detailed data (like the ones that we use) are shared in real-time and such transparency should minimize any biased behavior. Thus, in other social and economic settings with less public attention, one may expect a higher bias.

1.2 Data and empirical strategy

In this section, we first present our dataset, which has been compiled from several sources. Second, we describe our core empirical model and the variables we used.

1.2.1 Data sources

Our main dataset covers the universe of men's football matches of the top five European leagues (England, France, Germany, Italy, Spain). In a season, each team plays every other team twice, once at home and once as visitor. Over the period of 10 seasons (from 2011/12 to 2020/21), we have $N = 18,118$ matches.[12]

Such a dataset has several advantages. Multiple leagues allow filtering out possible country specific rules and customs and lead to high external validity, and a large coverage is also necessary to power our identification. Likewise, detailed and long time series information is necessary to examine referees' careers.

The main dataset is an event-by-event level description of every game, collected from whosco red.com. This records each and every action happening on the pitch. Each event has a type and a timestamp (at the second level). Event types include: pass, ball recovery, foul, tackle, throw-in, free-kick, yellow and red card, substitution, penalties, shot, goal, and corner. In a typical game, an event happens once in every 3.6 seconds: there are 1432 events recorded in the regular playtime of 90 minutes, and 90.1 events during stoppage time. Where relevant, the dataset also contains the location of the event on the pitch (in terms of x and y coordinates).

The second set of data is at the game level. It includes the venue of the game, attendance in the stadium, the result (goals by home and away teams), referee name, date, and time.

The third set of data concerns referees' experience in terms of the number of games they refereed in domestic and European competitions. The data have been collected from soccerway.com.

Furthermore, we used complementary data from a variety of sources. Information on stadiums comes from transfermarkt.com as well as from Wikipedia pages of teams. We used Deloitte Football Money League to identify top clubs.[13] We collected data to estimate the squad value of each team for every season, using historical player valuations from transfermarkt.com. Finally, for each league and season we downloaded the clubs' Élő rating score at the start of the season from elorating.com.

[12] There are 20 teams in a league (18 in Germany), $10 \times 4 \times 20 \times 19 \ + 10 \times 1 \times 18 \times 17 = 18260$. Due to the Covid-19 pandemic, the season 2019/20 of the French Ligue 1 finished early, with only 279 out of 380 games played. Due to data coverage deficiencies, we lost 41 games.

[13] See: Deloitte Football Money League Wikipedia page. The 20 teams include 6 English, 4 Italian, 4 German, 3 French, and 3 Spanish teams.

1.2.2 Empirical strategy

Our main outcome variable of interest, *Stoppage_time*, is the length of playtime (measured in seconds) beyond the regular time (90:00 minutes). We measure this as the timestamp of the game end in the event data. This is set to indicate the time on the clock (beyond 90 minutes) when the referee blows the final whistle. Note also that due to technical reasons, a typical gap of 2-3 seconds arises between the clock shown for the actual whistle and the end of the match timestamp in our data. However, as shown in Appendix A.1.2, this is orthogonal to the regular time standing and will hence not affect our results.[14]

The measure of *Stoppage_time* can (and often does) slightly differ from the expected stoppage time indicated by the referee at the end of the regular time, as referees may adjust the indicated time based on events like fouls and substitutions that happen during the stoppage time. Again, this variable is measured in seconds, allowing us to measure changes at a high level of precision.

We follow Garicano et al. (2005), and focus only on the cleanest comparison to study referees' favoritism bias: looking at matches where the goal difference at the end of regular time (90:00) is exactly one goal. Our key independent variable is an indicator variable, $Home\_lose$, as we compare the length of stoppage time in games when the home team is losing by one goal ($Home\_lose = 1$) with games where it is winning by one goal ($Home\_lose = 0$). For each game, events are aggregated at the first half, the second half, as well as the first and second stoppage time periods.

The average stoppage time is 253 seconds, ranging from 3 to 660 seconds. The home team wins slightly more games than the away team (57% compared to 43%), consistent with the well-documented general home advantage in sport (Jamieson, 2010). For a broader review of descriptive statistics, see Table A.2 in the Appendix. This difference between average stoppage time by the home team losing or winning, however, could be confounded by a variety of factors, such as injuries correlated with both stoppage time and the result, or differences in the playing style of the home team. To partial these out, we estimate the following model with OLS:

$$Stoppage\_time_{h,a,s} = \beta Home\ lose_{h,a,s} + \gamma Controls_{h,a,s} + \theta_l + \eta_h + \epsilon_{h,a,s}, \tag{1.1}$$

where our unit of observation is a single game played between home team h, and away team a in season s in league l. As each team hosts every other team once in a season, the h, a dyad uniquely identifies a game in any season s. We use a rich set of control variables ($Controls_{h,a,s}$) as described below. Standard errors are clustered at the home team level.[15]

As control, we include the following variables. First and most importantly, we approximate the justifiable length of stoppage time by counting the time during which the ball is likely to have been out of play in the second half. Our granular event dataset introduced above allows us to calculate

[14] We thank an anonymous referee for pointing out this discrepancy.
[15] Standard errors with the alternative home-away level clusters are slightly smaller.

the measure of *Wasted time* as a sum of seconds between two consecutive events if the first event is a foul, a card, a ball picked up by the keeper, or a goal; or if the second event is a corner, a throw-in, or a substitution. Thus, this variable captures all the events associated with time wasting, including those that determine the length of the stoppage time as per the Laws of The Game (see Section A.1 of Appendix). To make sure that the measure also captures longer interruptions of the games such as injuries or cooling breaks, we also add any interruption of the game that is longer than 30 seconds. In addition, to make sure that our results are not driven by extraordinary games with very long interruption (such as a serious injury), we exclude matches that fall within the top 5% of relevant matches in terms of the longest interruption.[16] In the second half, wasted time varies between 11 and 33 minutes, its average is 21 minutes (out of 45+4.5), equivalent to c. 60% effective playing time. Second, in addition to wasted time, we also control for the number of events associated with long stops: yellow or red cards, substitutions, fouls, and goals in the second half. It may matter, as referees may use heuristics, such as the number of these key events to decide stoppage time. Furthermore, these events are potentially confounding variables as they can be correlated with the goal difference as well, given that the playing style of the teams usually varies depending on winning or losing.

Third, a potential further confounding effect may be that instead of favoring the home team, referees may simply let the attacks started during the end of the stoppage time finish. It is a confounder because in general, the losing team is likely to play more offensively during stoppage time (as they need to score a goal), and the away team is more likely to lose (due to the home advantage in general). To control for this possibility, we generate a variable by taking into account the passes of the losing team during stoppage time, and take the average distance of these passes from the team's own goal line. This variable (called *Losing offensiveness*) is measured in units of distance from the team's own goal line, on a scale from 0 to 100.

Fourth, during the examined 10-year period, the video assistant referee (VAR) technology was introduced, and its use may affect both the stoppage time setting and the activity of players. Thus, we added a league-season level variable (VAR) indicating whether the technology was in use.[17]

Fifth, for each match we also control for the round of the season. This variable runs from 1 (the first match in the season for each team) to 38 (the last match in the season for each team in case of a league with 20 teams). In later rounds, there is more at stake in the game, which can affect both the referee's and team's behavior.

Finally, in line with earlier literature, we add league and home team fixed effects (θ_l and η_h, respectively) to capture the footballing style and quality of the team. This allows within-team comparisons of stoppage time conditional on end of regular time results. Our baseline specification does not include referee fixed effects, as we cannot rule out the possibility that the allocation of

[16] This step excludes matches where the game stopped for at least 266 seconds (almost 3 minutes).

[17] VAR has been in operation since season 2017/18 in Germany and Italy, since season 2018/19 in France and Spain, and since season 2019/20 in England. See more in Section A.1 in the Appendix.

referees is not random, and may be a part of the mechanism through which bias works. That said, all our results are robust to including referee fixed effects, as well as to including referee age (as a proxy for their experience) as a control variable.

1.3 Results

This section presents our empirical findings. We confirm the existence of a referee favoritism bias towards the home team, and show that this bias is not driven by the fans in the stadium. As a next step, we show that influential teams enjoy a larger favoritism bias from referees.

1.3.1 Home-team bias is still there, but not because of the crowd

In our dataset, 12.4 seconds is the raw difference between the additional stoppage time if the home team is losing by one goal and the additional time when it is winning by one goal. As shown in Table 1.1, once all control variables are added, this difference is marginally changed to 13.41.[18] Detailed regression results are shown in Table A.3 in the Appendix, revealing how each of the control variables is related to the length of stoppage time in our baseline specification with various fixed effects (Model 2 of Table 1.1). We see that stoppage time is indeed correlated with events in the game such as fouls or disciplinary actions. Furthermore, it shows that our baseline estimations are not sensitive to the choice of fixed effects, including fixed effects for the referees.

Due to the Covid-19 outbreak in Europe in spring 2020, practically every football league was suspended as of the second weekend of March. France closed the 2019/2020 season early, while other leagues resumed around May-June, with matches played behind closed doors. Closed door games continued in the 2020/21 season and in spring 2021, while partial opening meant that for 139 games stadiums were filled to an average capacity of 13%. Over these two years, about 2/3 of the games were played in fully or partially closed stadiums. (For details, see Appendix Table A.1 in the Appendix.) We created two indicator variables, $Closed = 1$ when attendance is zero, and $Covid = 1$ that also includes very low attendance games in 2021. As partial opening only affects 6.6% of the games, it will turn out to have very little impact.

To find out if the difference disappeared during closed games, we added an interaction term to equation (1.1) and estimated:

[18] Our results refer to the 2011-2021 period. Note that our estimated coefficient for Germany is 16 seconds, close to what was measured earlier by Dohmen (2008) but a magnitude smaller than the one in Garicano et al. (2005). For lack of available data from the period, it is difficult to make a direct comparison, but the difference in estimates is not driven by modelling choices: a replication of their core model (in their Table 2, column 4) offers a similar estimate to our favored specification (Appendix, Table A.4). It is possible that in the nineties there was less oversight of referee behavior in Spain.

Table 1.1: Presence of home bias, no crowd effect

	(1)	(2)	(3)	(4)
Home lose	12.40***	13.41***	13.81***	13.70***
	(2.07)	(1.65)	(1.60)	(1.61)
Home lose × Covid			-2.41	
			(5.19)	
Home lose × Closed				-0.99
				(5.08)
Controls	No	Yes	Yes	Yes
League FE	No	Yes	Yes	Yes
Home team FE	No	Yes	Yes	Yes
R^2	0.01	0.45	0.46	0.45
Observations	6667	6667	6667	6667

Note: Games with a single goal difference after regular time. Dependent variable is stoppage time in seconds. Controls include time with ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, round of season, whether VAR was used, and the average distance of the passes of the losing team from opponent's goal line in the extra time. In columns 3 and 4, all controls are also interacted with the Covid or the Closed dummy. Standard errors clustered at home team level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

$$Stoppage\_time_{h,a,s} = \beta_1 Home\ lose_{h,a,s} + \beta_2 Closed_{h,a,s} + \beta_3 Home\ lose_{h,a,s} \times Closed_{h,a,s} + \gamma Controls_{h,a,s} + \theta_l + \eta_h + \epsilon_{h,a,s}, \quad (1.2)$$

where our treatment indicator $Closed_{h,a,s}$ flags closed matches. An alternative is where instead, we flag all closed and partially open games ($Covid_{h,a,s}$). All control variables are also interacted with the *Closed* (or *Covid*) dummy to capture that without crowds players may behave differently or referees may take a different amount of time to make decisions.

Columns 3 and 4 in Table 1.1 show that the bias is unchanged whether the game is played in full or (mostly) empty stadiums: the estimated interaction terms are very close to zero. This result is robust to changing the mix of control variables (including adding referee fixed effects).[19]

The exogenous variation in crowd presence allowed us to test the hypothesis that the favoritism bias from referees favoring the home team is the consequence of the social pressure exerted by

[19] Another novelty during Covid was having cooling breaks in the summer games of 2020. We checked robustness by taking these games out, and it had no impact.

the fans of the home team in the stadium. Our results confirmed the existence of bias in referee decisions, but we can rule out that this is driven by the size of the crowd, as the bias remains unchanged even in the extreme case of closed stadiums. This is our first main result: any difference in the crowd size is very unlikely to be the mechanism behind social pressure, therefore, referees must be helping the home team for other reasons.

1.3.2 The home-team bias driven by influential teams

If favoritism bias from referees is not driven by crowds, there must be some other mechanism that leads to biased behavior.

In this section we investigate if the size of the home-team bias is correlated with team influence. Sporting success and financial clout allow teams to have influence: they attract players, fans, investment, or media interest. Influence will also provide teams with the capacity to exert social pressure in the form of the persuasion of independent agents.

Financial and sporting success are strongly correlated: wealthier teams will have better players and will win more often. In our baseline specification we use the league table ranking as it is a well-defined order of team success. In our data, wealth and quality measures are highly correlated with a correlation coefficient between 57% and 92%. In a robustness check, we use replicate results with other metrics of influence.

Ranking is defined as the end-of-season position of the team in the league table, the lower the better (1 is the title winner, 18 or 20 is the last team). The final ranking of a team at the end of a season is close to its expected average ranking throughout the season, and it may be easily compared across leagues and seasons.

To uncover the relationship between stoppage time and ranking, we first estimate a model with only the football rule controls (such as time with the ball out of play, number of cards, substitutions), as described in Section 1.3.1. Then, we compute the difference between predicted and actual stoppage time. This deviation, $\widehat{Bias}$ from equation (1.3), is the measure of the unexplained difference.

$$\widehat{Bias} = Stoppage\_time_{h,a,s} - \left(\hat{\gamma} Controls_{h,a,s} + \hat{\theta}_l + \hat{\eta}_h\right) \tag{1.3}$$

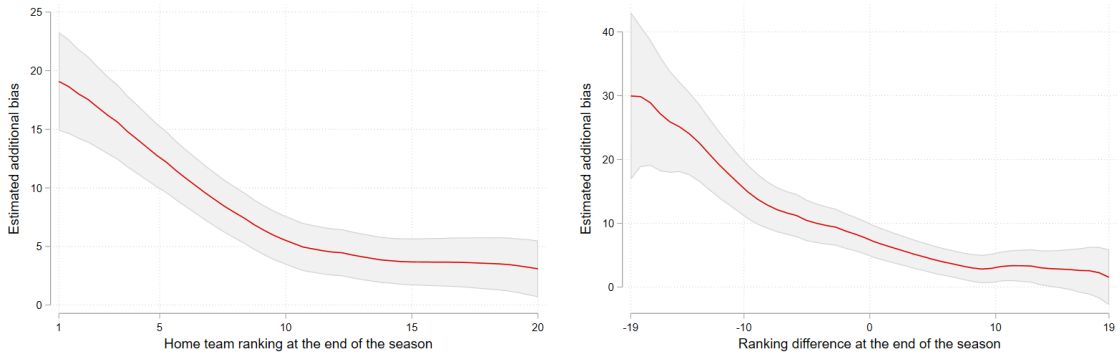
In the second step, using local polynomial smoothing regressions, we plot this bias against heterogeneity by the ranking of the home team. In Figure 1.1a, we see a fairly strong pattern with the higher-ranked teams enjoying a greater home-team bias. In Figure 1.1b, we see a similar pattern with the difference plotted against bias: when the top team plays against the lowest ranked one, the gap is 30 seconds, but it goes down to 0 when a minnow plays at home against a top team.[20]

Both graphs suggest that influence in terms of ranking may be non-linear. This may stem from

[20] Alternatively, confounders may be partialled out of rank as well, only to result in a very similar graph.

 DOI: *DOI: 10.14754/CEU.2024.07*

Figure 1.1: Heterogeneity of home-team bias by team ranking



(a) Home team ranking (b) Difference between home and away team ranking

Note: Local polynomial smoothing and a 95% confidence interval. Bias measured in seconds, rank between 1 (best) and 20 (worst); 1-18 for Germany. Predicted stoppage time is the residual from a regression of stoppage time on football rule controls (time with ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, round of season, whether VAR was used, losing offensiveness), all interacted with the *Closed* dummy.

an important rule: the top six teams will play in the rewarding European championships in the following season. Thus, we will also consider a binary variable, *Top6*, to measure influence by being inside or outside the top six teams per league.

We estimate this heterogeneity in two ways. First, we investigate the difference along team influence (here: ranking) with β_3 in equation (1.4) measuring the heterogeneity in home-team bias. The $Home\_ranking_{h,s}$ variable may be estimated in a linear form or with a binary Top6 variable.

$$\begin{aligned} Stoppage\_time_{h,a,s} = {} & \beta_1 Home\_lose_{h,a,s} \\ & + \beta_2 Home\_ranking_{h,s} + \beta_3 Home\_lose_{h,a,s} \times Home\_ranking_{h,s} \\ & + \gamma Controls_{h,a,s} + \theta_l + \eta_{hometeam_h} + \epsilon_{h,a,s}, \end{aligned} \tag{1.4}$$

Second, an alternative model uses the difference between home and away team ranking, with β_3 measuring the heterogeneity in home-team bias in terms of the difference between ranking. The $Home\_ranking\_difference_{h,s}$ variable may be estimated in a linear form or with a binary variable, where high difference is defined as a gap greater than -10. In both models note that home team fixed effects allow within team comparisons of different opponents.

$$
\begin{aligned}
Stoppage\_time_{h,a,s} = \beta_1 Home\_lose_{h,a,s} \\
+ \beta_2 Home\_ranking\_difference_{h,a,s} \\
+ \beta_3 Home\_lose_{h,a,s} \times Home\_ranking\_difference_{h,a,s} \\
+ \gamma Controls_{h,a,s} + \theta_l + \eta_h + \epsilon_{h,a,s},
\end{aligned}
\tag{1.5}
$$

The results of both models are presented in Table 1.2. In the simplest setup with the Top6 indicator in column (1) of Table 1.2, we see that influential teams enjoy a home-team bias that is more than twice the one for the rest (10.9 vs. 23 seconds). Looking at the home team's ranking in a linear way in Column (2), we see $22.08 - 1 \times 0.75 = 21.33$ seconds for the top team, reduced by 0.75 seconds per rank, shrinking to 7 seconds ($22.08 - 20 \times 0.75$) for the lowest ranked one.

Next, we add the difference between teams in terms of ranking from the perspective of the home team. The difference thus ranges between 19 (when the lowest ranked team plays at home against the best one) and -19 (when the top team hosts the bottom team).

In the binary setting (Column 3), we once again find that when there is a sizeable rank gap in favor of the home team, the bias is twice the size compared to a case with no difference (12.25 vs 12.25+17.35=29.6 seconds). In the linear model (Column 4), for equally ranked teams the bias is 14.29 seconds and the slope is -0.8. So the lowest ranked team actually has a -1-second (i.e., negative) bias when playing against the top team, while the top team enjoys a 29.5-second benefit against the lowest one.

These results suggest that referees have a home-team bias that is substantially higher for more influential (top-ranked) teams, especially when they play against less influential (low-ranked) ones. The results are rather stable across leagues as shown by Figure A.4 in the Appendix.

As noted earlier, influence captures aspects of financial and sporting success, and ranking is not the only way to measure it. We considered three alternatives.

First, perceived quality may be better captured by the so-called Élő rating of the teams. The Élő rating system, originating from chess, is based on past performance of the teams. The relative rating of two teams is designed to capture the expected outcome of the game.

Second, monetary wealth may be better proxied by the estimated squad value of each team at the start of the season. This is based on adding up individual player values for squads and using values from Transfermarkt.[21]

Third, another monetary measure is revenues generated by the club. Top revenues over 10 years is an indicator of being among the the 20 teams that have generated the highest average revenue over 10 years. It is a binary variable by design.

[21]For example, in the 2018/19 season, Arsenal, an English Premier League team, is valued at 659 million euros making it the 6th most valuable team (while the team holds the 5th position in the points table). https://www.transfermarkt.com/premier-league/startseite/wettbewerb/GB1/plus/?saison_id=2018.

Table 1.2: Regressions indicating home-team bias heterogeneity

	(1)	(2)	(3)	(4)
Home lose	10.86***	22.08***	12.25***	14.29***
	(1.74)	(3.51)	(1.65)	(1.71)
Home lose × Home Top 6	12.09***			
	(3.05)			
Home lose × Home rank		-0.75***		
		(0.27)		
Home lose × Home-away rank diff ≤ -10			17.35***	
			(5.74)	
Home lose × Home-away rank diff				-0.80***
				(0.20)
Controls	Yes	Yes	Yes	Yes
League FE	Yes	Yes	Yes	Yes
Home team FE	Yes	Yes	Yes	Yes
R^2	0.46	0.46	0.46	0.46
Observations	6667	6667	6667	6667

Note: Games with a single goal difference after regular time. Dependent variable is stoppage time in seconds. Controls include time with ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, round of season, whether VAR was used, and the average distance of the passes of the losing team from the opponent's goal line in the stoppage time. Control variables are also interacted with the *Closed* dummy. Standard errors clustered at home team level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Table 1.3 below we reproduced key results with alternative influential team definitions. In Model (1), *Home favorite* dummy is defined as the home team being one of the 20 richest teams as per Deloitte Football Money League. In Model (2), matches where home minus away Élő rating differences are in the top quartile within the given season and league are defined as games with home favorite. Model (3) defines Home favorite games as those where the difference between the squad value of the home and away team belongs to the top quartile within the given season and league. Our findings are robust to applying any of these definitions instead of our core metric.

To summarize, we find that influential (top-ranked) teams benefit substantially more from referee decisions. When playing at home, the top teams seem to create an environment that makes referees more inclined to help. It is the influence of the host team organization and not the crowd size that affects referee behavior.

Table 1.3: Robustness home-team bias heterogeneity

	(1)	(2)	(3)
Home lose	12.35***	10.55***	11.03***
	(1.74)	(1.75)	(1.69)
Home favorite	-2.84	-7.08***	-2.68
	(4.03)	(2.20)	(2.30)
Home lose × Home favorite	10.69**	13.41***	12.62***
	(4.32)	(4.01)	(3.93)
Controls	Yes	Yes	Yes
League FE	Yes	Yes	Yes
Home team FE	Yes	Yes	Yes
R^2	0.46	0.46	0.46
Favorite definition	20 richest	ELO diff	Squad value diff
Observations	6667	6667	6667

Note: Games with a single goal difference after regular time. Dependent variable is stoppage time in seconds. Controls include time with ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, round of season, whether VAR was used, and the average distance of the passes of the losing team from opponent's goal line in the stoppage time. Control variables are also interacted with the *closed* dummy. Standard errors clustered at home team level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Our findings may explain earlier evidence in the literature suggesting crowd and stadium characteristics affecting referee decisions. Team influence is correlated with many observable characteristics: stadium attendance and capacity or the type of stadium teams have.

Importantly, the variation in attendance and in the attendance-to-capacity ratio in particular are both strongly correlated with the popularity of both home and away teams. In our data, 75% of the variation in the attendance rate, an earlier measure of the crowd effect (Garicano et al., 2005) is explained by the home team, the league, the season, and the popularity of the away team. As such, no proper inference can be made on the basis of the variation in absolute or relative attendance (see Table A.5 in the Appendix).

Similarly, for running tracks (Dohmen and Sauermann, 2016), the top 6 teams in Germany have no tracks in their stadium but some of the rest (like Hertha Berlin, or Nürnberg) do. Hence, crowd influence loses its explanatory power once confounders are taken into consideration.[22]

[22]Close to fans is defined as not having a track or a sizable gap between the pitch and fans. An example for a stadium with tracks is Hertha Berlin for a covered gap is West Ham's London Stadium.

1.3.3 Is there corruption? Referee career in Europe

Finally, we investigate if career concerns may motivate referees to help influential teams. The specific question we address here is whether more biased referees are more likely to get work in the UEFA Champions League and in the UEFA Europa League games. Refereeing in these European games is a pinnacle of a referee's career, and the number of games they worked at is a key career success metric. Influential teams – especially those with regular presence in these competitions also having positions on UEFA boards – can (albeit rather informally) block unwanted referees and possibly promote preferred ones.[23]

To analyze careers, we aggregated game information into an unbalanced referee-season panel ($N = 1,148$) of referees ($N = 233$) and seasons ($N = 10$, $s = 2011 - 12 \dots 2020 - 21$). On average, we observe a referee for 4.9 seasons in their national leagues working at 70 games, 33 of which have a one-goal difference after 90 minutes.

We keep only referee-season pairs where we observe at least one 0:1 and 1:0 result after regular time and we are left with 179 referees and $N = 711$ observations. This dataset is merged with information on referee work during the same period in either UEFA leagues. We also have personal information on referees (age), as well as their past work in our sample (the number of games refereed in previous years).

For each referee-season pair, we define average favoritism bias from referees in two ways. First, for referee r in season s, $Bias_{r,s}$ is the average difference between the stoppage time when the home team is losing vs. winning. Second, $Bias\_pred_{r,s}$ is the average deviation from the predicted stoppage time for referee r when the home team is losing vs. winning, based on our model (1.3).

To analyze the relationship between European career and bias, we first estimate a cross-sectional linear probability regression, with the dependent variable, $Euro_{r,s}$ taking up 1 when referee r worked at least a single game at either competitions in season s and 0 otherwise.

Experience is a key potential confounder in this model, and hence we add age (linearly) (Age) and experience in terms of the number of seasons ($Experience$), as well as the number of domestic league games refereed (Dom_games). All right-hand side variables refer to the $s-1$ season. We also add league (L) and season (S) dummies.

$$\begin{aligned} Pr(Euro_{r,s} = 1) = \alpha + \beta_1 Bias_{r,(s-1)} + \gamma_1 Experience_{r,(s-1)} \\ + \gamma_2 Dom\_games_{r,(s-1)} + \gamma_3 Age_{r,s} + L + S + \epsilon_{r,s} \end{aligned} \tag{1.6}$$

Alternatives include restricting the sample for the 7th national league season only for each referee (Columns 3,4), replacing bias with predicted bias (Column 2), and having the number of European games instead of a binary variable (Column 4). Finally, (in Column 5) we repeat the model of

[23] See a short discussion in A.1.5 in the Appendix.

Column 1, with a different bias definition. $bias\_influence_{r,s}$ is calculated only for influential (top 6 ranked) teams. There are far fewer observations here, so we only estimate a pooled OLS.

Table 1.4: Referees' favoritism bias and UEFA jobs

	(1) Binary All All	(2) Binary All All	(3) Binary 7th All	(4) Count 7th All	(5) Binary All Influential
Mean home bias	0.0001		0.0010	0.0026	-0.0003
	(0.0002)		(0.0009)	(0.0032)	(0.0005)
Mean pred. home bias		0.0002			
		(0.0002)			
Domestic games (N)	0.0292**	0.0292**	0.0419**	0.2529***	0.0142
	(0.0081)	(0.0082)	(0.0173)	(0.0571)	(0.0110)
Controls	Yes	Yes	Yes	Yes	Yes
League FE	Yes	Yes	No	No	Yes
Season FE	Yes	Yes	No	No	Yes
R^2	0.38	0.38	0.36	0.46	0.40
Observations	553	553	78	78	225

Note: Column 1,2,3,5: Dependent variable is binary, refereed in Europe, linear probability model. Column 4: Count of games in Europe, OLS. In columns 1-4, the bias is the difference in the average residual when the home team is losing vs. winning. Predicted bias in column (2) is based on 1.3. In column 5, it is the difference between influential and non-influential teams when losing at home. Robust standard errors (col 3,4), referee level clustered standard errors (col 1,2,5). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Results are presented in Table 1.4. We find no correlation between bias and probability (or count) of refereeing European games. Thus, we see no correlation between any type of favoritism bias from referees and the likelihood of the referee working in the UEFA competitions. Referees do not (or cannot) expect any benefit in terms of European success in case they help teams that are likely to play in European competitions.

1.4 Conclusions

In this chapter, we extended and revisited earlier evidence of favoritism under social pressure, as measured by home-team favoring bias among European football referees. Using extensive data consisting of c. 7 thousand football matches from 10 seasons of the 5 top European leagues, we find that referees support the home team by allowing the game to last longer if it's losing. The point

estimate of this bias is 13 seconds or 5% of the extra time. The magnitude is comparable to earlier studies 20 years ago.

Our key contribution is understanding the source of social pressure pushing referees to make such decisions. Unlike previously assumed, the bias is not due to crowds as it exists even in the absence of home fans: it persisted when games were played in empty stadiums during the Covid pandemic.

Importantly, the bias is actually larger when it favors influential (top-ranked) teams and is especially large when an influential team is losing at home to a minnow. The point estimate of influential team bias is 24-30 seconds or 10% of the extra time.

To be certain that social pressure is in play rather than corruption, we showed that the favoritism bias from referees is, however, uncorrelated with career benefits.

The favoritism bias is quite sizeable in the sense that for more influential teams the 30-second extra represents an almost 10% longer stoppage time. We can translate it into financial benefit. The team that is losing at 90:00 equalizes or even wins on 7% of the matches, so there is a 0.7% chance that the home bias would change the points outcome of the game, on average by 1.3 points (sometimes lagging teams even win). With 38 games in a season, it adds up to almost 0.4 points. For a team like Manchester United, which spends on average 6.5 million USD per point, the bias would correspond to more than 2.5 million USD.[24]

Thus, instead of adherence to social pressure (crowds), we find evidence for adherence to host organizations, especially influential ones. So why do referees support these teams? After ruling out direct social pressure or corruption, we are left with a more nuanced explanation: it may be the consequence of an unconscious bias driven by a persuasive home team organization. People are often uncertain about their decisions and this is when this bias kicks in: favoring or being more willing to err on the side of a persuasive group. For instance, referees may make errors in not compensating enough for wasted time. Such an error may go unpunished if it hurts a small team but may get massive coverage in the media when a top team suffers.

This chapter has shown that it is hard to wipe out favoritism. Our findings show that there could be multiple sources of social pressure – in our case, influential organizations beyond the crowds. In the past decades, football organizations have (correctly) focused on taming crowds and punishing aggressive behavior. However, the power of some clubs may have even increased, with a global audience, more money, and higher stakes. Understanding the multiple sources of social pressure can help understand why crowd control is not enough to make referees cut favoritism.

While we were able to establish that home team bias is stronger for influential teams, there is room for future research in understanding how exactly influential organizations use their star power to exert social pressure. For instance, the analysis may be extended for influential start players, to test the extent to which their presence on the field exert pressure on the referees.

[24]Calculated based on the estimated salary costs and transfer fees collected from spotrac.com.

 DOI: *DOI: 10.14754/CEU.2024.07*

Furthermore, the results suggest that an influential team bias may exist on its own, independently from which team plays home or away. As the chapter's main focus is the decomposition of the home team bias, this question is out of the chapter's scope and left for future research.

DOI: *DOI: 10.14754/CEU.2024.07*

2 Chapter 2: The Xerox Effect: Communication Technology and Political Actions in Autocracy

Joint work with Arieda Muço

2.1 Introduction

> "... Our most successful program [in opening up the Soviet System] was providing Xerox machines."
>
> George Soros[25]

Throughout history, autocracies have been the dominant form of government in human history and still rule in more than a quarter of the countries worldwide. Yet over time, a significant shift towards democratic systems has emerged. As of 2017, democracies constitute 57% of countries worldwide, while autocracies and hybrid systems continue to exist in significant numbers.[26] While traditional media has been shown to be a key instrument in promoting electoral accountability in democracies (Snyder Jr and Strömberg, 2010; Strömberg, 2004) and a facilitator in the rise and consolidation of autocratic regimes (Adena et al., 2015), we know little about the role that media and other forms of information technologies play in ending autocratic regimes. Do information technologies help to promote democratic values and contribute to the fall of autocratic regimes? Do democratic values, in turn, promote growth and development?

We address these questions through an historical lenses: the dissemination of an innovative technology, photocopiers, in mid 1980s in Communist Hungary. The introduction of these machines happens at a point in time when only 12 photocopy machines existed in the country, all of which were securely locked away (Dányi, 2006; Nóvé, 1999) The spread of Xerox photocopy machines, in a country with limited freedom of speech and ideological monopoly of the state party was comparable to the Gutenberg press.[27]

The import of the machines was initiated in 1985 within what became eventually broadly known as the Xerox program. The program was one of several initiatives launched by the Hungarian Academy of Sciences - Soros Foundation, which was established in 1984 by Hungarian-born American businessman George Soros. This foundation was anchored to the Hungarian Academy of Sciences and supported various other programs such as scholarships and book imports.

The Xerox program specifically focused on equipping libraries, universities, research institutes, and other public institutions with cutting-edge Xerox photocopy machines. By 1989, approximately four hundred machines had been distributed throughout Hungarian cities, with over half of them

[25] Quote from the book "Opening up the Soviet System".
[26] Pew Research Center.
[27] Personal communication with Laszlo Bruszt, October 18, 2021.

located in Budapest, the capital. Soros, along with other politically active individuals in Hungary during that period, have widely recognized the importation and distribution of photocopiers - a technology previously not traded in the country - as one of the foundation's most successful initiatives (Kaufman, 2002; Soros, 1990).

To answer our main research question causally, we need two key elements: first, the random allocation of photocopy machines, and second, observable measures of political attitudes towards democratization and regime change in locations where these machines were distributed. To tackle the first, our analysis will leverage the variations in geography and time, as well as the unique historical circumstances surrounding the adoption of Xerox machines in Hungary. Regarding the second, we will use the 1989 referendum on democratic transition, organized by the political opposition, as a proxy. Our study then proceeds to examine the subsequent political developments, starting with the landmark 1990 general parliamentary elections and extending our analysis to more recent political outcomes, including the 2016 elections. This approach allows us to assess both the short- and long-term impacts of the technological intervention.

We find that in areas equipped with the new technology, voters are more likely to support proposals related to the details of the democratic transition in the 1989 referendum than areas without machines. Moreover, these areas are also more likely to participate in elections and exhibit decreased support for communist ideology as proxied by the vote for former communist parties. We also present suggestive evidence indicating that the presence of these machines positively influences neighboring municipalities, leading to greater participation in elections and stronger support for pro-democratic values.

Furthermore, using an event study approach, we examine the impact of these machines on various factors such as library users per capita, library stock and borrowed items. Our findings indicate that while there is no significant change in the library stock, the allocation of machines results in a decrease in the number of library users per capita, whereas the number of borrowed items per capita increases.

Overall, our results indicate that the dispersion of new technology facilitates the spread of information and helps with the spread affects both democratic values and entrepreneurial mindset. There are two potential mechanisms behind these effects. Similar to more modern information and communication technologies that have been extensively researched in the past years (e.g. Enikolopov et al., 2020; Manacorda and Tesei, 2020), the photocopy technology may impact political outcomes through enhanced information or enhanced coordination. The former reflects the possibility that the photocopy machines helped individuals either spread or acquire information that was not widely available before. The coordination channel, on the other hand, suggests that the technology helps individuals synchronize their actions. Anecdotal evidence and the data available suggest that the machines, imported with the initial aim to promote research and development were used to facilitate the circulation of educational content and access to previously restricted literature and

used to multiply samizdat journals (Dányi, 2006).

This chapter is related to several strands of the literature. First, it is related to the literature on the effect of information and communication technologies on political outcomes. Previous literature has documented that modern communication technologies (such as broadband networks in general or social media) have an impact on protest participation and mobilization (Acemoglu et al., 2018; Campante et al., 2017; Enikolopov et al., 2020; Kern, 2011; Manacorda and Tesei, 2020; Steinert-Threlkeld et al., 2015), economic growth (Czernich et al., 2011), and voter turnout (Campante et al., 2017; Falck et al., 2014). We focus on turnouts and voting behavior, such as anti-communist votes and referenda. We find that both turnout and democratic values are positively affected in the treated and neighboring areas.

Moreover, contemporary settings do not allow disentangling the demand and supply effects of political information. In these settings, both government and citizens can use traditional and social media platforms for communication purposes. In our setting, however, the supply of government propaganda through Xerox machines is non-existent. While the Hungarian government had tight control of all other media sources, including newspapers, television, and radio, the use of Xerox machines was outside of the government's grip.

Finally, this chapter is linked to the work of Dittmar (2011) who documents that European cities where printing presses were established in the 1400s grew faster than otherwise similar cities. Complementing this work, we focus on firm entrance and growth, rather than population growth, and find that entrepreneurial activity booms in areas where "independent press" is established.

2.2 Background

2.2.1 Political landscape

Shortly after World War II, until around 1947, Hungary was a pluralist parliamentary democracy with free general elections. The 1945 elections were held shortly after the end of World War II, and it was won by Independent Smallholders Party (*Fuggetlen Kisgazdapart*), by a huge margin, getting 57% of the votes. The Hungarian Communist Party (*Magyar Kommunista Part*), the predecessor of the state party between 1949 and 1989, got 17% of the votes.

Two years later, in 1947, the elections were held under much less democratic circumstances (Földes and Hubai, 1999).[28] By 1949, the country became a one-party system governed by the Hungarian Socialists Workers' Party, under the hegemony of the Soviet Union. Throughout the years of the communist regime, parliamentary elections were held regularly, but these did not have a real stake: in each district, only one state party candidate ran for office, and participation in

[28]The 1947 elections became infamously called the "blue-ballot" elections later: the Hungarian Communist Party rigged the elections by exploiting a loophole in the legislation that allowed individuals to vote away from their permanent address. This resulted in activists of the communist party casting multiple votes across the country on election day.

elections was strongly recommended: those who did not turn out could potentially be harassed by the state administration (Hubai, 2001b).

In the 1970s, in part as a consequence of the revolution in 1956, Hungary had a special status within the Soviet Bloc. Living standards were comparatively higher than those of other communist countries, and people enjoyed a higher degree of freedom in their everyday lives (Romsics, 2003). Despite the relatively soft regime throughout the 1970s and 1980s, the ideological monopoly of the state party was sustained and enforced. The state party enjoyed an information monopoly: every published material, including books, journals, and newspapers, was controlled by the government, and there was no room to convey alternative messages to the wider public (Haraszti, 2000). Science, media, art, and literature, were heavily impacted by censorship. Printing and distributing any kind of writing could not take place without previous approval.

The ideological monopoly started to be challenged around the 80s by a group of intellectuals, members of the democratic opposition. Besides some illegal self-education circles, the Budapest-based "flying universities", held in private flats with 100-200 participants, the most significant achievement was the ability to break free from the "intellectual ghetto" that dissent voices had been exiled for decades before (Bozóki, 1994). The intellectuals started to produce and distribute uncensored and illegal written material (books, journals, petitions, among other things) to a wider audience. The most famous platform for this was the flat of Laszlo Rajk Jr. The so-called *Rajk-boutique* operated once a day in the evening. People had the opportunity to make copies from the items available there, for which they had to pay the cost of multiplication.

The members of the democratic opposition managed to launch a few journals. The most prestigious, Beszelo, published 28 issues (samizdat) between 1981 and 1989 when it was legalized.[29] Through these platforms, a secondary public had evolved, where dissent voices, often promoting democratic ideas, could reach a few thousand people. The newly emerged secondary public was facing a significant technology barrier; the multiplication of the documents was a slow and costly process. For this purpose, the opposition used the few available stencil machines, mostly illegally smuggled to the country in parts (Hodosán, 2004). Technology for multiplying documents was not available to the wider public in this period; no more than a dozen photocopy machines were present in the country. The machines available were located at the highest levels of state administration, all under strict security surveillance (Nóvé, 1999).

The establishment of the democratic republic was declared in October 1989. In November 1989, a four-part referendum was held on the democratic transition. The referendum, broadly known as 'Four Yes Referendum', had four proposals, each concerning the democratic transition of the country. The first question asked whether the new president of Hungary should be elected after the first (democratic) elections, the second whether organizations related to the incumbent communist

[29]Samizdat means self-publication. Whereas Soviet and Czech samizdat books were published in single copies or in minimal numbers, the situation in Hungary and the German Democratic Republic was somewhat better Zaslavskaya (2013).

 DOI: *DOI: 10.14754/CEU.2024.07*

state party should be banned from workplaces, the third whether the state party should account for its owned and managed properties, and the fourth whether Workers' Militia should be dissolved.[30] The overall turnout was 58%, so the referendum was declared valid. All proposals passed. The first proposal one by a small margin (50.1%), the other three by around 95% of the votes. In Spring 1990, Hungary held free parliamentary elections.[31]

On May 2 1989, Hungarian soldiers began to dismantle the barbed wire and electric fencing along Hungary's border with neighboring Austria. The crumbling of communism in Hungary, marked a significant event amidst a wave of revolutions in Central and Eastern Europe. Despite tight security, Hungarian authorities attempted to prevent the influx of refugees. The dismantling of the 240-kilometer electric fence along the Hungary-Austria border served as a pivotal crack in the "Iron Curtain" that had divided Europe for over four decades since the end of World War II. This event set in motion a series of events, ultimately leading to the fall of the Berlin Wall.

2.2.2 The Hungarian Academy of Science/George Soros Foundation

In 1984, Hungarian-born American businessman George Soros established his foundation in Hungary. The Hungarian government's decision to allow the operations of Soros Foundation in Hungary can be mainly attributed to economic motivations. Facing financial constraints, the government recognized the potential influx of resources that the foundation could bring into the country. Soros, in his book 'Opening the Soviet System,' highlights this economic angle, noting that while the Ministry of Economics was in favor of the foundation's presence due to its potential economic benefits, there was reluctance and skepticism among other groups, including artists.

The establishment of the foundation involved extensive negotiation, and Soros insisted on his right to withdraw all his money and cancel his entire commitment if a pattern of political interference emerged in the selection of the grantees (Kaufman, 2002; Soros, 1990). A key condition set by the authorities was that the foundation, and Soros by extension, should abstain from supporting political activities. In return, through a gentlemen's agreement, Soros received assurances from the government that individuals with a 'troubled' political past would not be denied access to the foundation's resources (Kaufman, 2002). To integrate the foundation within the existing political structure, the Communist Party demanded that it be anchored within the Hungarian Academy of Sciences, a prestigious intellectual entity under the regime. Additionally, the authorities stipulated a change in the foundation's governance structure: instead of six board members as Soros envisioned, they insisted on two chairmen, proposing the secretary-general of the Academy of Sciences to co-chair alongside Soros (Dányi, 2006; Kaufman, 2002). Kalman Kulcsar was the member serving alongside with Miklos Vasarhelyi, as Soros's surrogate or personal representative on the board. The foundation was called, the Hungarian Academy of Science/George Soros Foundation.

[30] A paramilitary organization of the state party between 1956 and 1989.
[31] The elections were won by the leading conservative nationalist party.

Both board members had veto power – the rest of the committee, according to Soros, were independent-minded Hungarian intellectuals. However, neither Kulcsar nor Vasarhelyi exercised their veto power (Kaufman, 2002).

Soros himself was not given a free pass. In his book Opening the Soviet System, he recounts "I interviewed the candidate for the position of executive secretary put forward by the academy, I said to him, 'You have a tough job serving two masters.' 'Only two?' he replied, which I understood to imply that he also had to report to the security agencies."[32] Since the security apparatus insisted that the administration of the foundation and people who would process the applications were recruited from their core, a compromise was worked out. Just as there were two chairmen of the board, there were two executive directors supervising day-to-day work (Kaufman, 2002).

2.2.3 Activities of the foundation

The foundation provided grants, scholarships, travel stipends for Hungarian scholars to visit other institutions. One of the main activities was also to provide books, some of them restricted material, in libraries. While the propaganda apparatus of the Communist party attempted to impose a ban on publicity concerning the activities of the foundation, the foundation placed advertisements in newspapers and published annual reports. Economic Weekly (HVG) would make publication the activities of the foundation and the public became aware of the activities of the foundation only gradually (Soros, 1990).

Eventually, the foundation decided to import photocopy machines inspired by the challenge it was to copy material in Hungary. Before the arrival of the Xerox machines the country, few machines were in use, all under lock (Dányi, 2006). Photocopying was only allowed with prior official approval; in case of acceptance, people would leave their material for photocopy and would only be able to pick it up later. Even if the material was approved, it would take weeks to receive it. The emergence of hundreds of new Xerox machines equaled a figurative and literal information revolution and a clear metaphor for open society (Dányi, 2006; Kaufman, 2002; Nóvé, 1999). George Soros himself has broadly acknowledged that importing and distributing photocopy machines was one of the most successful initiatives of the foundation (Soros, 1990).

2.2.4 Allocation of the machines

The call for application for the Xerox program was first advertised in Spring 1985, and it was prolonged for almost a decade afterward. The machines, state-of-the-art technology, were imported from the US, and arrived in Western Europe by boat and were moved within the country by train after the initial landing in Budapest. The first machines made it into the country in the Summer

[32]Soros goes on saying "We had some very hard negotiations both before and after the signing of the agreement. The officials thought that they were dealing with a well-meaning expatriate, the proverbial American uncle, whom they could humor and take advantage of."

of 1985. Similar shipments continued for several years. The recipients of the machines received the machines at a discount price.

In the first two years of the program (1985 and 1986), the call for application was coordinated and disseminated by the Association of Hungarian Librarians (*Magyar Könyvtárosok Egyesülete, MKE*). The applicants were mostly public libraries, and libraries of universities and hospitals. The foundation was in charge of compiling the list of accepted applications; the list was then either modified or approved by the Board of Trustees (Nóvé, 1999). Contemporaneous archival documents of the foundation suggest that in these first years of the program, most applications were approved.[33]

In 1987, as people became more aware of the foundation and the Xerox machines, the foundation and the program came under scrutiny from the state party. Janos Kadar, the Hungarian prime minister, attempted to intervene and limit the foundation's activities, including the Xerox program (Kaufman, 2002). Although the attempt was not successful (in part due to George Soros personally endorsing the continuation of the activities), the programs of the foundation (including the Xerox program) became more politicized. Indeed, from 1987 onwards, civil and political organisations became frequent recipients of the machines (Nóvé, 1999).

Former members of the illegal democratic opposition acknowledge that these machines played a crucial role in allowing information to flow freely. For instance, Zsolt Enyedi, a political science professor and former opposition member, reflects on the transformative power of these machines: "Xerox broke my dissident career... By the time we put the instrument together, the first Xeroxed samizdat publications appeared. Home-made copying lost sense."[34]

In the totalitarian framework, where the unwritten rule is that anything not explicitly permitted is prohibited, the spread and use of the Xerox machines subtly challenged this principle by enabling unapproved information dissemination (Kaufman, 2002).

2.3 Data sources and summary statistics

Our empirical strategy relies on a comparison of geographic areas that had access to the photocopy technology with those that had no access at the municipality level. For this, we control for a wide set of covariates that capture the characteristics of the municipalities prior to the Xerox program. These covariates include official statistical data, census data, and election results. This Section describes the sources of these data and presents some summary statistics.

Documentation on the distribution of the Xerox machines in the country are publicly available in the annuals of the foundation. The documents contain the comprehensive list of institutions that

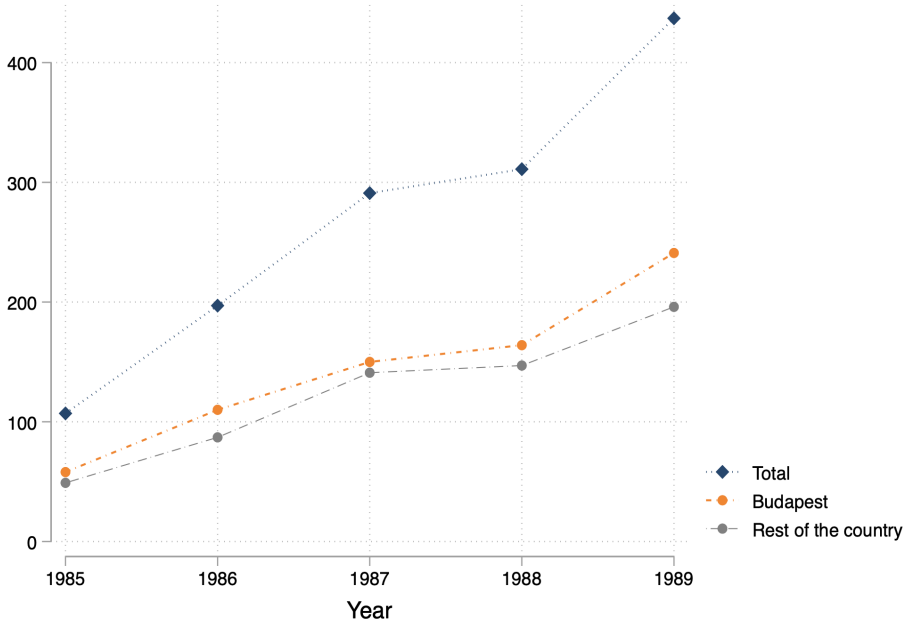
[33]The foundation's archival material is stored in the Vera and Donald Blinken Open Society Archives (OSA). In these archives, the approved and rejected applicants are available only from 1988 onwards, but not for the first years of the program.

[34]Zsolt Enyedi's recollection provides a personal perspective on the pivotal role of Xerox machines in changing the dynamics of information dissemination and contributing to the democratic opposition's efforts.

were granted with photocopy machines between 1985 and 1989, except for 1987. We complement this data source with the archival documentation of the Xerox Program, handled by the Vera and Donald Blinken Open Society Archives (OSA). The OSA archives allow us to get the list of grantees for the year 1987, not contained the annual publications. The detail in the data sources allows us to also extract the type of institutions granted with a machine and geolocate them via Google Places API.

Figure 2.1 plots in blue the stock of Xerox machines in the country for each year. The stock of machines in Budapest is shown in orange: in gray, instead, the rest of the country. As it can be clearly seen from the figure, the stock of machines distributed in Budapest is roughly equal to the stock of machines in the rest of the country. In 1985, the initial year of the program, there were distributed a total of 107 machines in the country, 58 of which in Budapest. The stock of machines in 1989 was of 437 machines, 241 of them were located in Budapest.

Figure 2.1: Stock of Xerox machines in the country



Notes: The stock of Xerox machines in the country for each year, in blue. In orange instead is shown the stock of machines in Budapest and in gray the rest of the country.

Access to the municipality level aggregate data based on the 1980 census on the demographic

and economic structure of Hungarian municipalities was provided by the Hungarian National Social Science Data Archive. Historical data, covering the 1880 census, were made available by GISta Hungarorum project. The 1880 census help us get an estimate of the population of the municipality. The latter is captured by the share of population engaged with associations. These associations in the late 19th century were advocacy groups formed mostly in the area of economic, political, cultural, or sports activities (Sebestény, 2003). We use the membership rate in these associations may be an indicator of the strength of the civil society in a given municipality.[35] We also use data on the capital of financial institutions, normalized by population. The decades preceding 1880 brought an unprecedented boom in newly established financial (mostly credit) institutions: by the year of the census we have data on, their network had covered the country (Tomka, 1996). Their capital per capita captures the economic development and financial awareness of the municipalities.

The data on turnout and vote shares from the general elections of 1945 and 1947 were generously provided by historian Laszlo Hubai. The election outcomes for the 1985 parliamentary elections in Hungarian counties, which we scanned and digitized, can be found in Hubai (2001a). The National Election Office of Hungary provided data on the 1989 referendum, the 1990 general elections, and the 2016 referendum. We digitized the hard copies of the Statistical Annuals of the Hungarian Central Statistical Office (KSH) from 1980 to 1994 to obtain data on library subscribers, stock, and borrowed items.

Table 2.1 presents mean comparisons for demographic, economic, and political characteristics of Hungarian municipalities from 1880 to 1980. Column (1) presents averages for all areas equipped with Xerox machines. Column (2) shows the averages for the municipalities that never received a machine but have a library or hospital located within them. It can be clearly observed that there are significant differences among areas across most socio-economic characteristics included in the table. Areas with machines were more populated in 1880 and 1980, had a higher literacy rate, and also a larger GDP per capita, among other characteristics.

The rest of our control variables are derived from the census carried out in 1980, four years before the foundation was established.[36] Besides population and density, we use information on the shares of population broken down by gender and occupation sector, as well as on the average housing space per capita.[37] Beyond those, we use information on the existence of a hospital within the municipality, as well as the stock and number of subscribers of the municipality library, normalized by population. There are considerable differences between treated and control groups in all these variables, except for electoral outcomes in 1945 and 1949 and library stock and subscribers. Treated municipalities are much more populated and also more dense. The most striking difference can be identified in the share of people working in agriculture; more than 40% in the control group, while only 13% among treated municipalities. All these disparities suggest a sharp urban-rural distinction

[35] These regional attitudes are widely considered to be highly persistent over time (North, 1994).
[36] 1980 is the last year prior to the Xerox program for which census data are available.
[37] Kern (2011) uses a similar set of covariates.

between areas with and without machines.

Table 2.1: Mean comparison

Variable	Control	Treated	Diff.
Population (1000), 1880	1.62	13.24	11.62***
Literacy rate (pct.), 1880	42.31	52.44	10.13***
Association members (pct.), 1880	0.81	10.71	9.90***
Capital of financial institutions per capita, 1880	0.84	9.06	8.22***
Turnout 1945	93.71	92.64	-1.06
Communist 1945	13.87	15.58	1.71
Population (1000), 1980	2.33	41.47	39.14***
Population density (pop/km2)	0.73	4.12	3.39***
% female	52.66	49.97	-2.69***
% active	43.82	49.16	5.33***
% industry	11.39	19.01	7.63***
% agriculture	40.36	12.82	-27.55***
% construction	3.13	3.75	0.63**
% services	12.12	19.74	7.62***
% higher education	1.57	4.93	3.36***
Housing space (m2) p.c.	205.79	83.16	-122.62***
Hospital (y/n)	0.04	0.81	0.77***
Stock of city library, per capita, 1980	4.10	3.74	-0.36
City library subscribers, per capita, 1980	17.69	17.22	-0.48
Observations	2,146	73	2,219

Notes: Budapest, and municipalities without hospital, university and public library are excluded from the sample.

However, political variables, such as turnout and vote share for the Communist Party in the last two free elections (specifically, in 1945 and 1947), do not significantly differ across areas. Similarly, the per capita stock of libraries and the number of city library subscribers are comparable across both the treatment and control groups.

We use the municipality level vote shares of the Hungarian Communist Party on these elections to capture regional attitudes towards (anti-)democratic ideas, as we do not find any significant difference between the control and the treated municipalities in 1945 and 1947.

2.4 Empirical strategy and results

Our main hypothesis is that the photocopy machines distributed across Hungary as part of the Xerox program affect political outcomes, and democratization of the country. To formally test for this hypothesis, we start estimating the following model:

$$Outcome_i = \alpha + \beta Treated_i + \gamma X_i + \tau_c + \epsilon_i, \tag{2.1}$$

where $Outcome_i$ is the outcome of interest in area i, measured after the treatment. Ideally this outcome would be change in regime, since change of regime happens at the country level, we will proxy for it by voting behavior in the municipality level.

$Treated_i$ represents either an indicator variable taking value one if the municipality i received a photocopy machine under the umbrella of the Xerox program. The vector X_i is a set of pre-treatment characteristics measured before 1980. τ_c are county fixed effects. ϵ_i is the random error term that accounts for unobserved factors affecting the outcome. Standard errors are clustered at the municipality level. Our parameter of interest is β which measures the impact of treatment, Xerox machines in the municipality, on a set of outcomes.

We use different outcome variables used to capture democratic attitudes. First, we use turnout and vote shares of different outcomes for the referendum held in 1989, we aggregate the vote for the 'Yes' questions. Then we move to the general elections held in 1990. These elections, held in two rounds in March and April 1990, were the first free and multi-party elections since 1947.

The Hungarian Democratic Forum (*Magyar Demokrata Forum*), a new party founded in 1987 and on the right of the political spectrum, won the elections receiving 23.9% of the votes. They also formed the government in a coalition with two smaller right-wing parties. The two legal successor parties of the communist state party, the Hungarian Socialist Party (*Magyar Szocialista Part*) and the Hungarian Socialist Workers' Party (*Magyar Szocialista Munkaspart*) received 10.2% and 2.7% of the votes, respectively. We construct a measure of post-communist vote shares by taking the sum of these two parties' vote shares in each municipality.

To assess whether the effect of distributed machines persisted in the long run, we analyze the outcome of a referendum held in 2016, three decades after the Xerox program. The referendum, referred as "quota referendum" was initiated by the government to reject the European Union's migrant relocation plans.[38] Opposition called for either invalid voting or a boycott of the referendum; as a result, the vote share of *No* votes (the government's preferred option, rejecting the migration quotas) overwhelmingly dominated (98.3%) among the votes cast – the referendum was declared invalid as the share of valid votes was well below the 50% required by the law (44% turnout, with a 6.1% share of blank or invalid votes).

We use the vote share of *No* votes (pro-government option) for each municipality. Since the campaign of the government personally targeted George Soros personally, we expect areas equipped with photocopy machines granted by the foundation to be less prone to succumb to the anti immigration and anti-Soros propaganda.

[38]Hungary was one of the affected countries during the European migrant crisis. On 17 June 2015, Viktor Orban's Fidesz government announced the construction of a 175-kilometre-long fence along its southern border with Serbia. See Reuters article.

We estimate an OLS regression of Equation 2.1 using voting behavior of individuals as outcome variables and the set of covariates presented in Table 2.1. Treatment is defined as an indicator variable which takes value one if the municipality received a photocopy machine at any point in time. That is, we compare areas with and without access to the technology.[39] As shown in Table 2.2, we cannot document significant short-term effect for any outcome except for the Post-Communist vote share in 1990.

The OLS results suggest that while Xerox machines do not affect the 'Yes' referendum in 1989, or participation rate in the first free elections, areas with Xerox machines are more likely to vote Post-Communist parties and tend to support the 'anti-Soros' campaign in the 2016 referendum.

Table 2.2: OLS estimation

	Yes 1989	Turnout 1990	Postcom 1990	No 2016
Treated	0.127	0.897	2.536***	1.744**
	(0.328)	(1.024)	(0.682)	(0.886)
Controls	Yes	Yes	Yes	Yes
Observations	1148	1867	1867	1888
R^2	0.422	0.460	0.328	0.313

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

2.4.1 Instrumental variables approach

Our parameter of interest resulting from estimating equation (2.1) is likely to be biased due to the unobserved characteristics that affect both technology adoption and vote pro-democratic values: it cannot be ruled out that the Xerox machines were allocated to politically more active regions. We address this issue using an instrumental variable approach, leverage the fact that the early allocation of the machines was not influenced by politics: as described in Section 2.2.2, the politicizing of the Xerox program began only in 1987. Therefore, we assume that the machines allocated in 1985 and 1986 are not correlated with unobserved regional political characteristics. This assumption is supported by the fact that in these years, the machines were given to public institutions; by contrast, in later years, machines were also granted to nascent civil organisations and political parties.

While we cannot formally test the validity of our instrument, we provide some evidence for it. For this we use the 1985 parliamentary elections. On 8 June 1985, Hungary conducted its parliamentary

[39] Alternatively, treatment can be defined as the number of machines granted to the municipality (that is, measuring the treatment effect on the intensive rather than the extensive margins). Furthermore, the treatment effect can be separated by year in which the municipality received the first machine. The results of these estimates are shown in Appendix B.1.

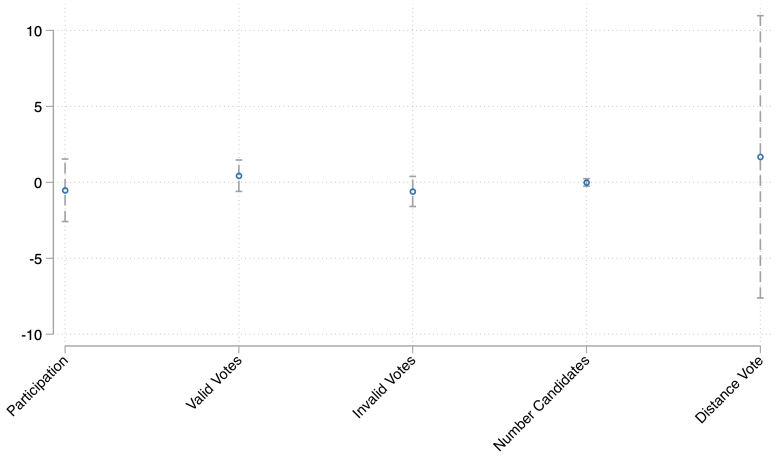
elections. In these elections, for the fist time under the communist regime, opposition candidates were also eligible to run for seats. Therefore, unlike previous elections held during the Soviet era (which were no more than a formality as only the state party's candidate ran for each seat), the 1985 elections had real political stake and voters had the possibility to express their discontent with the state party (Racz, 1987). The state party secured 288 out of the 387 seats in the election.

We estimate the following specification:

$$Outcome_{i,1985} = \alpha + \beta Treated_i + \gamma pop_i + \tau_c + \epsilon_i, \tag{2.2}$$

where $Outcome_{i,1985}$ is the outcome of interest such as participation rate, valid or invalid votes, number of candidates, and difference in the vote share between the first candidate with the second. $Treated_i$ is an indicator taking value one if a machine is allocated in the area in 1985 or 1986 and zero otherwise. pop_i is a control for the population, measured in 1980, and τ_c are county fixed effects. ϵ_i is the usual error term.

Figure 2.2: Elections 1985: Early treated



Notes: The figure plots coefficient estimates from a regression of the the allocation of the machines in 1985 or 1986 on each outcome. The treatment indicator takes value one if the municipality receives a machine in 1985 or 1986 and zero otherwise. Budapest is excluded, but results are qualitatively similar if Budapest is included in the sample.

The figure above plots the coefficient estimate from each separate regression and indicates that the allocation of machines in 1985 or 1986, does not affect 1985 electoral outcomes. The balanced pre-electoral outcomes suggest potential similar political attitude of treatment and control areas

before the distribution of the photocopy machines.

Hence, we proceed with the instrumental variable approach where the early treatment is used as an instrument for treatment. Results from the IV estimation are shown in Table 2.3. All our coefficients are in line with the expected sign suggesting that in areas where machines were allocated, voters were more likely to vote for democratic transition in the 1989 referendum, were more likely to participate in the 1989 elections, and less likely to vote for anti-immigration policies in 2016.

Table 2.3: IV estimation

	(1) Yes 1989	(2) Turnout 1990	(3) Postcom 1990	(4) No 2016
Treated	4.341***	4.333***	1.536	-7.766***
	(0.396)	(0.989)	(1.143)	(0.747)
Observations	1,148	1,867	1,867	1,888
R-squared	0.016	0.003	0.002	0.003

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.4.2 Propensity score matching

We also carry out a nearest-neighbor propensity score matching on Hungarian municipalities that could potentially have been exposed to treatment. To construct the control group we consider municipalities that had public libraries, universities, or hospitals.

Hence, we disregard municipalities where there was no hospital, public library, or university established in 1985, in the starting year of the Xerox program. We also exclude Budapest from our matching exercise as an obvious outlier (both in terms of covariates and treatment) without any meaningful potential matches from the untreated group.

The propensity score is estimated with a logit model. The outcome of interest is an indicator variable that takes value one if a photocopy machine was allocated in the area and zero otherwise. We include a set of pre-determined covariates as explanatory variables. Based on the estimated propensity scores, a nearest-neighbor matching with replacement is carried out. That is, for each treated unit, we assign a unit from the control group, the one with the closest estimated propensity score. An untreated unit can potentially be matched to more than one treated unit. As the last step, the average treatment effect (ATE) is estimated as the mean of the outcome differences between treated and matched control units.

As shown before in Table 2.1 of Section 2.3, there are considerable differences between treated and control municipalities in most covariates used for the propensity score estimation: the entities that received the photocopy machines tend to be larger, more industrialized, and more developed

in general. The difference is also captured in their legal status: the vast majority of treated entities were cities. Since the legal status of the Hungarian municipalities had undergone major restructuring during the years of the Xerox program, we do not put any limits on our sample based on municipality classification.

The covariates have been selected such that the set controls for the historical heterogeneity across regions in terms of political attitudes, economic circumstances, and demographic characteristics, among other things. These attributes seem to be highly persistent over time: even as early as 1880, both literacy and association membership rates show significant differences between control and treated regions, and so does the capital of financial institutions per capita.

To control for pre-treatment political attitudes, we use the information on the general elections held in 1945 and 1947, the last multiparty elections for over four decades. In particular, besides turnout (a natural proxy for political activism), we compare the vote share of the Hungarian Communist Party (*Magyar Kommunista Part*). We do not observe any significant differences here. This suggests that, despite the huge gaps in their levels of economic development, the control and treated regions are comparable at the pre-treatment period in terms of political attitudes and preferences, and they did not seem to self-select into treatment on the basis of this. Moreover, as shown in Appendix B.2, the balance diagnostics indicate that the matching exercise can sufficiently balance the treated and the control groups.

The results of the propensity score matching estimation are presented in Table 2.4. We can observe highly significant average treatment effects for short term outcomes (specifications in columns (1)-(4)), all with the expected signs. The estimation suggests that having a photocopy machine in a given municipality increases the share of supportive votes in 1989 referendum and turnout in the 1990 general elections by 2.2 and 5.4 percentage points, respectively. Given that the overall turnout was 65% in the general elections, this is a considerably large effect. Besides turnout, the voting patterns also differ in treated municipalities: residents in these municipalities tended to favor less the successor parties of the communist state party (by 4.3 percentage points on average). The results also show that the effect is persistent: the estimation in column (5) suggests that in treated municipalities, individuals are less succumb to the anti-Soros state propaganda, supporting it by 4 percentage point less on average in the 2016 referendum.

Table 2.4: Propensity score matching

	(1) Yes 1989	(2) Turnout 1990	(3) Postcom 1990	(4) No 2016
ATE	2.219***	5.417***	-4.320***	-3.965***
	(0.165)	(0.338)	(0.418)	(0.398)
Observations	1149	1872	1872	1893

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A robustness check of the matching estimations is carried out in Appendix B.3, in which we repeat the propensity score matching exercise without limiting the sample to municipalities with either public library, hospital, or university. Those results show that while the short-term estimations are highly robust across these two specifications, the estimation regarding the 2016 referendum changes significantly and is of the opposite sign. This suggests that the hypothesis that the photocopy machines may potentially have had a long-lived consequence for people's political attitude and, in particular, voting behavior cannot be supported by reliable empirical evidence as the results are highly sensitive even to minor changes in sample construction.

We can see that these results are in line with the findings from the IV-strategy. The main difference is the vote share on post-Communist parties that in this case is negative and significant.

2.4.3 Placebos

To further substantiate that the results are not driven by unobserved characteristics, we conduct a placebo experiment by estimating with OLS the effect of the early-allocated Xerox machines on political outcomes that took place *before* the machines were allocated. Namely, we focus on the voting behavior in the 1945 and 1947 elections that took place well before the start of the Xerox program. During this period, the Xerox machines were not available anywhere in the country, hence, we do not expect them to affect predetermined voting outcomes. In light of this, we expect that the treatment variable is uncorrelated with these outcome variables. The opposite would indicate that unobserved factors are present that affect both the treatment variable and the political outcomes at the municipality level. Such finding would undermine the credibility of the main results as well.

Results from our placebo analysis are shown in Table 2.5. The results suggests that the future allocation of Xerox machines granted in 1985 and 1986 is not associated with the electoral outcomes in 1945 and 1947, therefore unobservable characteristics of municipalities, which could potentially be correlated with political preferences of the electorate, on the one hand, and with the machine availability, on the other, are unlikely to explain our main findings.

Table 2.5: Placebo estimation of the effect of treatment

	Turnout 1945	Communist 1945	Turnout 1947	Communist 1947
Treated	0.781	0.525	-0.010	-0.271
	(1.118)	(1.648)	(0.028)	(1.567)
Controls	Yes	Yes	Yes	Yes
Observations	1888	1888	1887	1892
R^2	0.021	0.260	0.047	0.184

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.4.4 Spillover effects in nearby areas

The analysis so far focused on the impact of the Xerox machines within the municipality in which the machine was allocated into. However, in reality, the boundaries between neighbouring municipalities are often blurred: many people commute, on a daily or occasional basis, either for work or leisure purposes. This is particularly true for cities and other larger municipalities with public libraries and universities (which were the primary recipients of the Xerox machines): residents of nearby municipalities are likely to be regular visitors of these institutions and are as such exposed to the effect of the treatment effect. We therefore hypothesize that the treatment has an effect on neighboring municipalities as well.

To test this hypothesis, we estimate the following linear equation:

$$Outcome_i = \alpha + \gamma Distance_i + \gamma X_i + \epsilon_i, \tag{2.3}$$

where $Outcome_i$ represents one of the five outcome variables described before, measured in municipality i; $Distance_i$ stands for the air distance of municipality i from the closest available photocopy machine; vector X_i is composed by the same set of covariates presented in Table 2.1. The spillover effect is measured by the parameter γ. Unlike our main estimates described before, this spillover estimates also includes municipalities without public libraries and universities, as these municipalities are also likely to be exposed to the spillover effect. The treated municipalities are excluded, however.

The estimation of the spillover effects are presented in Table 2.6. The results suggests that proximity to a Xerox machine is associated with more democratic attitudes: more supportive votes in the 1989 referendum, and lower post-communist vote share in the 1990 elections. Turnout in 1990, on the other hand, seems not to spill over as expected, as proximity to machines is associated with lower rather than higher turnout. Furthermore, unlike the effect on treated municipalities, the spillover effect is not persistent, as distance to the machines has no impact on the voting behavior

Table 2.6: OLS estimation of spillover effects on untreated municipalities

	Yes 1989	Turnout 1990	Postcom 1990	No 2016
Distance (km)	-0.024**	0.048**	0.034**	-0.017
	(0.010)	(0.019)	(0.016)	(0.018)
Controls	Yes	Yes	Yes	Yes
Observations	1176	2506	2506	2530
R^2	0.410	0.429	0.312	0.294

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in the 2016 referendum.

2.4.5 Mechanisms

To understand the mechanisms behind our findings we rely again on the institutional settings. According to (Kaufman, 2002), the greatest benefit of the copier program was the positive momentum it generated for the activities of foundation. Support for projects and travel stipends were not unknown in Hungary, but in the past they were awarded mostly on the basis of political patronage. Now, in a revolutionary departure, ads appeared in academic journals and leaflets were distributed at university dormitories inviting broad participation in open competitions based on merit. Anyone could submit outlines for projects that would be considered by the board. Scholarship applications for study abroad were welcome, as were request for support of ongoing research in Hungary.

The Xerox program outperformed the book project in terms of success, allowing environments to replicate a diverse array of materials at will, ranging from research papers and love letters to financial documents, as well as political and religious texts, and even, presumably, censored materials (Kaufman, 2002).

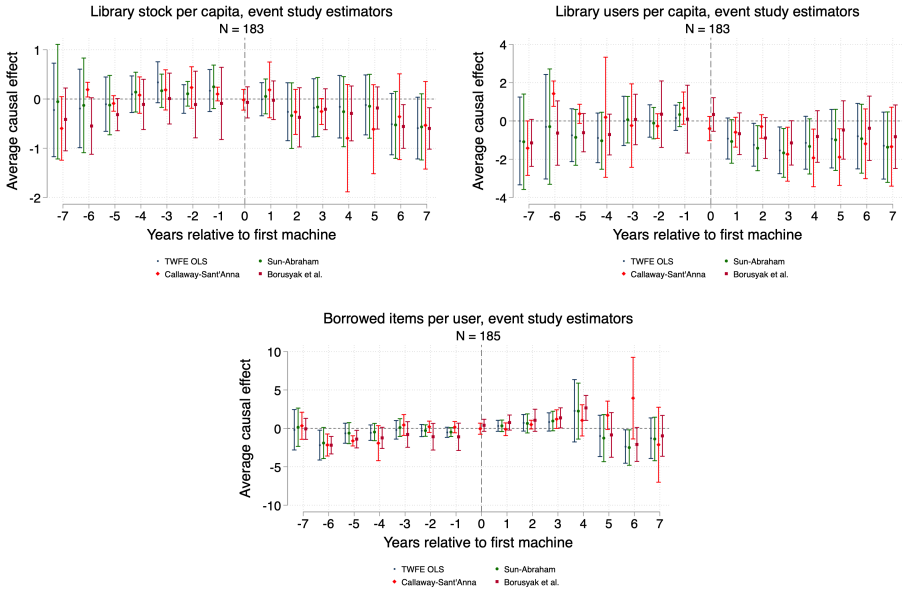
However, it is possible that the books project, rather than the photocopiers, are the driving force behind these findings. In order to rule out alternative mechanisms, such as the effect operating through other programs of the foundation such as book provision, we use data on library stock per capita, library users per capita, and borrowed items per user. We use the staggered treatment timing, and the availability of the data at the yearly level to estimate the effect of the treatment on the outcomes of interests. More specifically, we estimate the following equation:

$$Outcome_{it} = \alpha + \sum_{k} \gamma_k D_{ik} + m_i + \delta_t + \epsilon_{it}, \tag{2.4}$$

Where $Outcome_{it}$ is the outcome of interest of municipality i at time t. D_{ik} are indicator variables tracking the years, t, that immediately precede and follow the allocation of a machine

 DOI: *DOI: 10.14754/CEU.2024.07*

Figure 2.3: Library Stock, Borrowed Books, and Subscribers



Notes: The top left panel plots the stock of books per capita. The top right panel plots the library users per capita. The bottom plot instead plots borrowed items per user.

in the municipality i. m and δ are municipality and year fixed effects, respectively. ϵ_{it} is the idiosyncratic error term. We are interested on the coefficients on the event time indicator denoted as γ_k. Results are shown in Figure 2.3.

Our event study design indicates an absence of impact from the machines on the library's stock, suggesting that the increased book importation from the foundation is unlikely to explain our observations. Furthermore, we observe a decline in library users per capita, which may imply a shift in preference towards university libraries. Conversely, there is an upward trend in the amount of material borrowed per capita. This trend aligns with the hypothesis that the availability of Xerox machines led to an increased propensity for users to access and circulate photocopied materials. The robustness of these findings is shown across various estimation methods.

2.5 Conclusion

In this chapter, leveraging unique and newly digitized and georeferenced data on the spread of photocopy machines across Hungary in the late 1980s, we provide evidence on how novel information technology shapes the socioeconomic landscape and democratic attitudes at the municipal level.

 DOI: *DOI: 10.14754/CEU.2024.07*

Consistent with existing anecdotal narratives, our findings indicate that photocopy machines played a pivotal role in Hungary's democratization process. They did so by facilitating a decentralized dissemination of free ideas, contributing to a more open and informed society. The setting seems to suggest that the new technology enabled individuals not only to access information but also to actively participate in the democratic process.

These insights bring to light the complex and layered impacts of technological advancements on societal dynamics. They highlight the importance of examining and understanding the broader implications of such technologies in shaping socio-political landscapes. As we continue to navigate an era of rapid technological change, our study serves as a reminder of the critical need to explore and comprehend the far-reaching consequences of these innovations in redefining societal norms and behaviors.

DOI: *DOI: 10.14754/CEU.2024.07*

3 Chapter 3: The Causes and Consequences of International Student Migration and Mobility

3.1 Introduction

The increasing importance of international student mobility presents both opportunities and challenges to countries and educational institutions. This chapter addresses a key aspect of this phenomenon: the post-graduation return decision of international students, a factor that is essential for drawing consequences on the economic and social impacts of international student migration. It is well established that international students tend to flow from developing to developed regions. If a sufficiently high number of these students return to their home country after graduation, they contribute to the global knowledge diffusion and economic convergence. Low return rates, on the other hand, impose significant harms on the source countries in the form of brain drain.

The extent of post-graduation return rate is an empirical question in essence. In absence of robust empirical evidence, one can only rely on macro-level data and theoretical models, which offer only suggestive insights, at best. As this chapter will show, such analysis, based on aggregated country- and university-level data, has limited capability to explore the factors influencing the return decision of international students. Therefore, to bridge this gap, as a next step, this chapter introduces a novel contribution: a detailed, self-designed web-based survey aimed at directly capturing the motivations and intentions of international students concerning their post-graduation plans.

Before presenting the details and the results of the survey, the chapter first discusses the suggestive evidence on return rates that can be drawn from the combination of theoretical models and existing country-and university level data. A theoretical model that focuses on the human capital aspect of migration predicts relatively higher return rates; a model that puts student mobility in a migration model framework, in turn, predicts relatively lower return rates. As the two models posit different driving factors behind the students' mobility decision (i.e., the reason they choose to study abroad), the validity of them can be assessed empirically based on student flow data. This can therefore implicitly suggest their return behavior. I carry out these empirical tests focusing on Europe, using both country- and university-level data in a (semi-)gravity model framework.

These macro-level analyses suggest that the student flow patterns are slightly more consistent with the predictions of the human capital model, thus lending more credibility to this model. As the human capital predicts that students choose destination on the basis of university quality rather than post-graduation opportunities, this implies a relatively high post-graduation return rate.

These findings are then compared with the results of the self-designed survey, based on the responses of 600 students of the Central European University. These responses show a nuanced picture on return rates: a slight majority indicates that they are rather unlikely to return their

home country in the near future. Among the factors that affect the return decision, economic and political conditions and employment prospects play a great role, while those who plan to return do so mostly for family reasons.

The policy implications of these findings are rarely in the forefront of discourse. International student mobility, although arguably a net positive phenomenon, may impose significant burden on the source countries and hence contribute to the brain drain and the concentration of human capital in a few developed regions. This has serious negative impact on the competitiveness of the affected regions. In order to prevent this, source countries should aim to sustain a sufficiently high level of return rate. Based on the survey results presented in this chapter, this seems to be achievable through improving economic and employment conditions in the country. Before that, however, a first step should be to establish a system that follows the career path of foreign students which could support data-driven policy making.

3.2 Background

Improving the international mobility of higher education students has been a major policy priority in the agenda of the European Union.[40] Efforts to internationalize higher education date back to more than two decades now in Europe: the Bologna Process started in 1999 with a goal to ensure standardization and transferability of higher education qualifications across countries. Looking back now at the overall figures, the project can be considered successful: the number of foreign students enrolled in a higher education institution different from their home country has been in a constant rise over the last twenty years, expanding from 2 million in 1998 to 6.4 million in 2021 worldwide (UNESCO, 2023).

A student's mobility decision has key consequences not only for the individual itself, but also for both the source and the host country. Mobility is generally considered beneficial for host countries, as it usually imposes import of talented young people who can contribute to innovation-driven growth in knowledge-based economies and are also key to the university finances through the collected tuition fees. In line with this, many host countries design policies to attract foreign students; some even define targets regarding the number of incoming foreign students. The United Kingdom, for example, has set to reach 600,000 foreign students by 2030 in 2019, which they have already achieved by 2022[41]; in Germany, 350,000 foreign students were targeted by 2020 in 2013, a figure which was already reached in 2017; the same happened to Canada, where the 2022 target of 450,000 incoming students has been exceeded in 2017.[42] Policies to attract foreign students usually include recruitment campaigns abroad, lowered entry barriers to student migration, extended visa

[40] See, e.g., the "Mobility and cooperation" goals of the European Commission.

[41] https://www.gov.uk/government/publications/international-education-strategy-2023-update/international-education-strategy-2023-progress-update

[42] *Wissenschaft Weltoffen 2019*, https://nbn-resolving.org/urn:nbn:de:0168-ssoar-66584-6

and work permit opportunities after graduation, and specific scholarship programs.

On the other hand, in several large host countries the drawbacks of large-scale student inflow have become more prevalent recently. In the Netherlands, for example, it is widely held that the large number of foreign students arriving to the country have contributed to the imminent housing crisis. There are ongoing policy plans to limit the inflow of foreign students.[43] Nevertheless, the net effect of the arrival of international students is arguably positive, but ultimately it depends on the return rate of the graduates: if only a small proportion of international graduates remain in the country, the benefits may be limited.

The effect from the perspective of the source countries is much more ambiguous. When their talented citizens go abroad for study, they are likely to acquire valuable knowledge and skills, which will be beneficial for the country – provided that they return at some point after graduation at all. In this case, the source country benefits from the mobile students' contribution to knowledge absorption, research and development activities.

If, on the other hand, they permanently remain abroad after graduation, it will impose a considerable loss of human capital on the source regions. This brain drain mechanism would lead to lower productive potential and accelerated population ageing, as also recognized by the European Commission (Lutz et al., 2019). This process may enhance disparities in the level of development between regions, and as such, eventually work against the European Union's cohesion policy goals. Indeed, brain drain is identified as the main factor inhibiting the convergence of emerging European economies and their escape from the middle-income trap (Mátyás et al., 2022).

The welfare effect of international student mobility therefore ultimately depends on the post-graduation return rate of the students, both for the source and the host country. Source countries are interested in *high* return rates, as they benefit if the student returns and utilizes their absorbed knowledge and skills there. Lower return rate, on the other hand, implies a permanent decrease in the country's stock of human capital and, when large scale, severe brain drain. Host countries, on the other hand, profit from *low* return – that is, talented foreigners remaining in the country and driving the country's economy. If, however, host countries provide the education for students who only come for the purpose of the study and leave straight after graduation, the benefit is limited to the period of study (e.g., through tuition fees and foreign students' local spending during their studies).[44]

Thus, the return of international graduates is a question of high policy relevance, with crucial consequences. Economic models, combined with observable patterns in macro-level data can provide reasonable predictions, at best. Consistent direct data on return rates are not tracked. Even the basic statistics on return rate are scarce, while systematic information on the determinants of the

[43] https://www.government.nl/latest/news/2023/04/21/steps-to-improve-the-management-of-incoming-international-students

[44] This is particularly true for countries that host a large number of foreign students which causes social tensions (see the Dutch example above). These tensions are only exacerbated when students just come and go.

return decision is practically nonexistent.

3.3 Theoretical considerations

In the economic literature, generally two broad theories are used to study the topic of international student mobility, each of which comes to different predictions and conclusions regarding the motives and consequences of foreign education decisions of individuals: the human capital model and the migration model. I briefly review these below, with a focus on the models' prediction on the post-graduation return behavior of international students.

The *human capital model*, originating from Becker (1964), considers education as an investment tool for an individual to maximize permanent income. In this framework, international migration takes place to acquire skills that are not available at one's home country. As the return of human capital investment is increasing with the quality of education, students who choose universities with this intention in mind consider the quality of university as a crucial factor. This model therefore postulates that students flow to countries with internationally recognized higher education institutions (and to high quality universities in particular) that provide portable skills. The students weigh the expected returns against the costs, which, beyond the costs of education, include the costs of moving to another country as well. The human capital model therefore predicts that students are likely to prefer the highest quality university that is available to them at the lowest possible cost. In this framework, student choose a destination *university* in the first place, and the host country itself plays a smaller role. Students who choose university based on human capital are therefore less likely to remain in the country in which they study and more likely to return to their home country (or settle down elsewhere) (Beine et al., 2014).

The *migration model*, on the other hand, considers student emigration as a tool for settling down permanently in the destination country. In this framework, students primarily choose university based on the destination country, and the quality of the university itself plays a smaller role. Migration takes place if the net benefits from switching country of residence are positive. As in many countries student visa is easier to acquire than a visa for permanent settlement, migrating as a student can reduce the cost of moving for someone who plans to settle down permanently in the destination country (Kaushal and Lanati, 2019; Rosenzweig, 2008). In this sense, higher education can be considered as a stepping stone for permanent settlement. Compared to those who choose university for human capital purposes, those who are planning to settle down permanently are more likely to move to longer distances, as for them the costs of migration are offset on the long term. The migration model predicts that students flow from countries with poorer economic circumstances, less employment opportunities, and less political stability, to more developed countries (i.e., with better performing economy, job market, and political stability). Therefore, in this framework these macroeconomic indicators are good predictors of the extent of student flow.

The two models therefore have different predictions about the post-graduation behaviour of international students: according to the human capital model, students are more likely to return to their home country, while the migration model predicts a lower return rate and a higher likelihood of permanent settlement in the host country. Although this is a key question with significant policy implications, in absence of data on post-graduation return behavior it is difficult to draw empirical conclusions. We can only use the insight of these models to deduce the return behavior: if we observe that the student flow patterns are consistent with the human capital model (i.e., that university quality rather than country factors play a great role), it may suggest that the return rate will be higher. If, on the other hand, student flows are more consistent with the migration model (that is, the macro-level factors of the origin and destination country are more important than university quality), it implies that return rates are likely to be lower, as these students seem to move with the purpose of permanent settlement in the destination country.

Empirically assessing the student flow patterns can help shed light on motivations of students to study abroad. These, in turn, can be considered as suggestive evidence for the validity of the human capital and the migration models. As discussed, these models have different predictions on the post-graduation return behaviour. Therefore, a student flow analysis can provide indirect, suggestive evidence on return rates.

In what follows, I carry out these analyses using country-level and university-level student flow data, which will then be compared with the responses to the original survey.

3.4 Country level analysis

3.4.1 Data

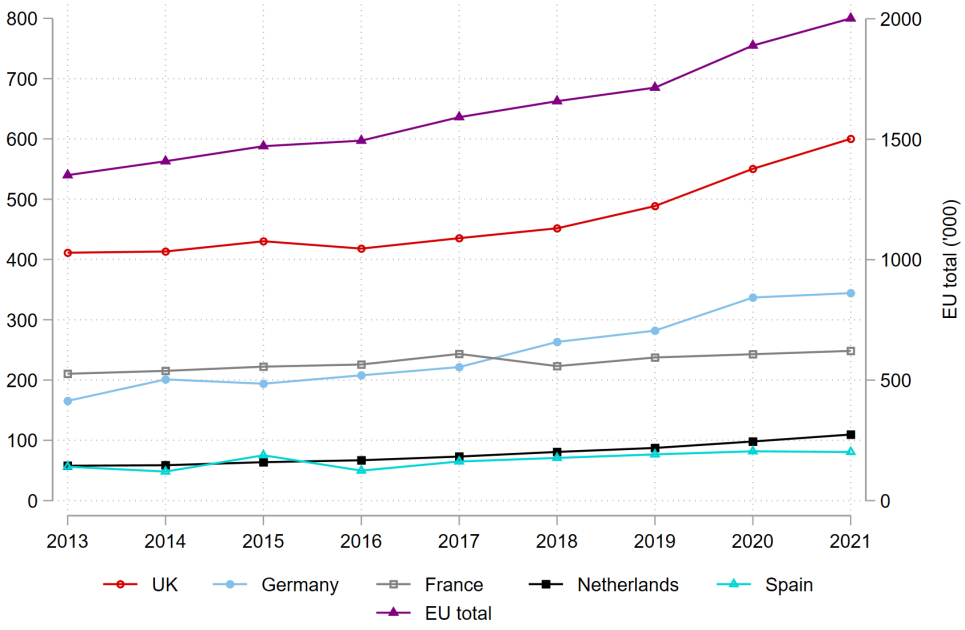
This section briefly describes the data sources used for the country level analysis, along with a brief description of the data. For assessing student flows across countries, I merge data from several sources. First, data on degree mobile students from abroad by country of origin, jointly administered by UNESCO, OECD, and Eurostat. This contains the number of foreign students arriving to EU countries annually, broken down by country of origin, gender, and education level. Second, the bilateral gravity dataset of CEPII (Conte et al., 2022), which gathers a set of macroeconomic and geographic factors of origin and destination countries at the yearly level. Third, the annual *Freedom in the World* dataset, maintained by American non-profit organization Freedom House, evaluating the political and civil rights in countries annually. Finally, to capture average university quality in each country, I calculate the average score of the universities listed in the world university rankings of Center for World University Rankings for each country and year.

The combined dataset covers 9 years between 2013 and 2021, 36 destination countries and 211 origin countries. That is, the dataset covers students arriving to the European Union (plus the United Kingdom), originating from practically all over the world.

 DOI: *DOI: 10.14754/CEU.2024.07*

The dataset allows one to establish some high-level facts. First, there has been a steady growth in the number of degree-mobile students studying in the EU, with the total figure rising from c. 1.4 million in 2013 to 2 million in 2021.[45] Despite some strong concerns expressed by experts around 2020, the Covid-19 pandemic did not seem to have any material medium-or long-term impact on mobility: as shown in Figure 3.1, there is no break in the upward trend in 2020 and 2021, the years of the pandemic. Second, mobile students are strongly concentrated in a few large host countries. The most popular destination country is by far the United Kingdom, which hosts 600,000 students as of 2021. The UK is followed by Germany, France, Netherlands, and Spain. These 5 largest host countries (as shown in Table 3.1) account for more than two-third (69%) of the total number of mobile students in the EU as of 2021. The source countries are also fairly concentrated, although to a lesser extent, with the 5 largest countries accounting for 31% of the total.

Figure 3.1: Degree-mobile students by host country in the EU, 2013-2021 ('000 students)



[45]International student mobility has two distinct forms. *Degree mobility* involves enrollment into a foreign university for a complete program, in the course of which a degree is earned abroad, while *credit mobility* refers to short study visits at a foreign institution; the credits taken during these visits are usually admitted by the sending institution. The latter does not fall within the scope of this chapter.

Table 3.1: Largest host and source countries, 2021

Top 5 host countries ('000 students)			**Top 5 sending countries ('000 students)**		
Rank	Country	Students	Rank	Country	Students
1.	United Kingdom	600.2	1.	China	240.2
2.	Germany	344.2	2.	India	142.6
3.	France	248.4	3.	Germany	101.4
4.	Netherlands	109.7	4.	France	68.8
5.	Spain	80.7	5.	Italy	67.1
Total		2,001.8	*Total*		2,001.8

3.4.2 Empirical strategy

For the country-level bilateral data, I estimate a gravity model. This model is standard for estimating trade flows between country pairs, but is equally well suited for assessing migration flows, including those of students.[46] The model setup is the following:

$$S_{ijt} = Y_{it}^{\alpha_1} Y_{jt}^{\alpha_2} D_{ijt}^{\alpha_3} \gamma_i \gamma_j \delta_t, \tag{3.1}$$

where indices i, j, and t stand for the host country, the source country, and year, respectively. S_{ijt} stands for the outcome variable (in this case, the number of students flowing from country i to country j in year t; Y is a vector of country-level factors; D is a vector of dyadic country factors (such as distance between countries i and j. The parameters γ_i, γ_j, and δ_t stand for host country, source country, and year fixed effects, respectively. Adding fixed effects to the model guarantees that the results are not driven by unobservable, time-invariant host or source country characteristics, nor by general time trends, as these are absorbed by the fixed effects. Thus, the parameters of interest (α_1, α_2, and α_3) are identified from the relative changes between country pairs from one year to another.

As for country-level characteristics (vectors Y_{it} and Y_{jt}), I control for the natural logarithm of GDP per capita for both countries, the democracy index for both countries (which are significant predictors of student flows according to the human capital model), and the university quality of the host country (which is a strong predictor according to the human capital model).[47] The dyadic factors (vector D_{it}) are the natural logarithm of the distance between the origin and the destination countries; and a set of binary variables indicating whether the origin and the destination countries (i) are contiguous, (ii) share a common official language, and (iii) have a common colonial history.

[46] For the microfoundations of the gravity model of migration, see e.g. Beine et al. (2016) and Ramos (2024).

[47] Controlling for the university quality of the country of *origin* would considerably reduce the number of observations, as this variable is only available for a small subset of developed countries.

I estimate the model with Poisson Pseudo-Maximum Likelihood (PPML), which is well suited for gravity models (Silva and Tenreyro, 2006). It is the standard method for estimating migration flows in particular (Ramos, 2024). It has several advantages over other methods, notably that the estimators are consistent even if the distribution is misspecified (so long as the conditional mean is correctly specified). Moreover, the method is also robust to overdispersion, which makes it widely applicable to modelling count data. The estimated parameters can be interpreted as elasticities for control variables included as natural logarithm, and as semi-elasticities for control variables included as levels.

3.4.3 Results

Table 3.2 shows the estimation results. First, Model (1) is estimated for all students, irrespective of the degree of studies. It shows that the most decisive factors in the student flows are geographic or historic. Namely, proximity of the countries, sharing a common language, or common colonial history all predict more intensive student flows. This result is consistent across all study levels (bachelor, masters, doctoral; Models (2), (3), and (4), respectively), except that contiguity of countries itself has no significant effect at the doctoral level.

Besides geographic factors, the development of the origin and host countries have no effect on the extent of student mobility for the students, except at the doctoral level; there is some suggestive evidence though that students tend to go study abroad from countries with lower democracy index to a larger extent (though this effect is only marginally significant for all students and the master students). The level of democracy in the host country has no impact, however. The effect of average university quality in the host country is only significantly positive in case of the doctoral students.

To test the extent to which the motivating factors of student migration differ from migration in general, I also ran the same estimation with the overall migrant flow between countries as the outcome variable. These results are shown in Model (5).[48] In case of these migrants, the geographic and historic determinants play the primary role, similarly to students. Besides these, not surprisingly, migration tends to flow towards countries with higher GDP per capita; in particular, a 1% higher GDP per capita implies a more than 2% increase in migration flow.

[48]This data was downloaded from Eurostat.

 DOI: *DOI: 10.14754/CEU.2024.07*

Table 3.2: Country-level gravity model to estimate student mobility

	(1) All students	(2) Bachelors	(3) Masters	(4) Doctoral	(5) All migrants
ln(GDP per capita origin)	-0.143	-0.238	-0.092	0.025	-0.217
	(0.141)	(0.152)	(0.168)	(0.127)	(0.155)
ln(GDP per capita destination)	0.614	0.215	0.717	2.335***	2.112***
	(0.485)	(0.468)	(0.638)	(0.437)	(0.451)
ln(distance origin-destination)	-1.107***	-1.172***	-1.014***	-0.920***	-1.125***
	(0.060)	(0.066)	(0.064)	(0.053)	(0.049)
Contiguity dummy	0.345***	0.619***	0.220**	-0.135	-0.018
	(0.093)	(0.099)	(0.100)	(0.084)	(0.083)
Common language dummy	1.599***	1.756***	1.505***	1.371***	0.819***
	(0.071)	(0.079)	(0.079)	(0.068)	(0.090)
Colonial history dummy	0.468***	0.402***	0.461***	0.775***	0.274***
	(0.069)	(0.076)	(0.074)	(0.069)	(0.090)
Democracy index origin	-0.011*	-0.009	-0.013*	-0.006	0.004
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Democracy index destination	-0.006	-0.020	0.011	-0.019	-0.009
	(0.015)	(0.017)	(0.015)	(0.013)	(0.014)
University quality destination	0.009	0.006	0.007	0.039***	-0.002
	(0.009)	(0.009)	(0.014)	(0.008)	(0.010)
Origin FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.83	0.80	0.84	0.83	0.85
Observations	36328	36328	36328	36328	24349

Note: Standard errors clustered at origin-destination country pair level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In summary, these results suggest that the motivations for student migration are generally different from other source of migration. Instead of the economic conditions in the host country (which are significant determinants for non-student migrants), students seem to focus on university quality in the host country (although the latter effect is not statistically significant). This does

not hold for doctoral students, however: they seem more similar to (non-student) migrants in that GDP per capita in the host country has practically the same (highly significant) effect on their host location choice. This is not surprising as doctoral student status is considered employment in some countries; also, doctoral students are probably more likely to remain in their host country after graduation. Unlike (non-student) migrants, for doctoral students host country average university quality seems also an important factor.

These results are similar to many others in the literature that, using country-level data, argue that the economic development of the host country is not the main driver of student flow. This is confirmed by papers assessing single destination countries, in particular for students arriving e.g. to Germany (Bessey, 2012), the United Kingdom (Ragot et al., 2017), and Italy (Beine et al., 2020). Bilateral country-level analyses also find that the quality of higher education is the main driver of student flow, as opposed to the income effect of the host country (Brezis and Soueri, 2011; Thissen and Ederveen, 2006; Van Bouwel and Veugelers, 2013).

Altogether, these findings indicate that students are less likely to act according to the *migration model*. This may be interpreted as a suggestive evidence for the predictions of the human capital model, as students are more likely to neglect their future potential income after graduation if they plan their stay abroad to be temporary. In this case, a significant share of students is assumed to return to their home country after graduation.

Controlling for the average university quality in each country does not take into account the variation of in the quality of universities within countries. To take a more thorough look at the role that the host universities (as opposed to the host countries) play in the destination choice of international students, the next section therefore turns to university level data.

3.5 University level analysis

3.5.1 Data

The university level analysis is mainly built on the dataset built under the European Tertiary Education Register (ETER) framework.[49] It is a comprehensive database of higher education institutions in Europe, covering c.3000 higher education institutions from 37 countries, currently covering 10 years between 2011 and 2020. The data is provided by the national statistical authorities. The register uses the UNESCO's, OECD's, and Eurostat's definition of statistics (and international mobility in particular), but at the institution as opposed to the country level (Lepori et al., 2023). It provides information on educational activities, research activities, expenditures, and staff of universities. Besides these, it also covers both degree and credit mobility, which is broken down according to the International Standard Classification of Education (ISCED): that is,

[49]The European Tertiary Education Register (ETER) is funded by the European Commission, under the contracts EAC-2013-0308, EAC-2015-0280, 934533-2017 A08-CH and EAC-2021-0170.

the number of mobile students is available separately for the bachelors, masters, and doctoral levels. However, this information is only available for a subset of the universities (c.1,300 universities at the bachelors and masters level; only c.750 at the doctoral level). There is no data on the country of origin of mobile students.

I collected annual university rankings by Times Higher Education (THE) from their website for years 2011-2020, and merged them with the ETER data applying fuzzy name matching. With this method, I could assign a ranking score (on a scale of 1-100) to c.1,300 university-year observations in the ETER data (which corresponds to c.15% of the observations with non-missing mobility information).[50] Finally, I added country-level yearly characteristics from the CEPII dataset (described above) for the country of each university.

The combined dataset contains 3,441 observations for which the relevant variables are all non-missing. This covers 812 universities from 15 different countries, between 2013 and 2020. As shown in Table 3.3, 4 out of the 5 largest host universities in the sample are from the UK, with the University of Vienna being the only exception with c.20k international students. Importantly, however, some large international hubs such as Dutch or French universities are not covered by the mobility section of ETER, which is an important limitation of this dataset.

Table 3.3: Largest host universities, 2020

Rank	University	Students	Share
1.	University College London (UK)	23,355	59%
2.	University of Vienna (AT)	20,705	27%
3.	The University of Manchester (UK)	15,585	43%
4.	The University of Edinburgh (UK)	15,550	46%
5.	King's College London (UK)	13,760	44%

3.5.2 Empirical strategy

As the university-level data is not broken down by the country of origin of foreign students, only a 'semi-gravity' model can be estimated, in the following form:

$$S_{uct} = X_{ut}^{\alpha_1} Y_{ct}^{\alpha_2} \gamma_c \delta_t, \tag{3.2}$$

where indices u, c, and t stand for host university, host country (in which the university is located), and year, respectively. S_{uct}, the outcome variable, is the flow of international students to the given

[50] Including this control variable in the model therefore substantially reduces the number of observations in the estimation. However, as shown in Annex C.1, the results are very similar when the ranking score is not included as a control variable.

 DOI: *DOI: 10.14754/CEU.2024.07*

university in the given year; X is a vector of university-level factors; and Y is a vector of country level factors. The parameters γ_c and δ_t stand for host country and year fixed effects, respectively. University fixed effects are not added to the model as the university-level control variables (budget, student fees, student to faculty ratio, and ranking score) vary relatively little from one year to another. This model therefore intends to better capture the factors that explain the variation in international student mobility across universities *within* a country.

3.5.3 Results

Table 3.4 shows the estimation results first for all students (Model 1), and then separately for bachelors (Model 2), masters (Model 3), and Doctoral (Model 4) students. The results indicate that considering all students altogether, the ranking score of the university does matter to international students, as higher ranked universities attract significantly more of them. The country level control variables (the natural logarithm of GDP per capita, and the democracy index of the country) are not significant, however, which confirms that students primarily choose universities and not host countries. The other university level control variables are also all statistically significant: the total budget of the university, the total funding from student fees, and the student-to-faculty ratio all have a positive effect on international student info (although the latter is only statistically significant at the 10% significance level).

These results are broadly consistent with the results of the country-level analysis. In particular, these strongly suggest that generally the host university matters to international students more than the host country. This can be considered as further suggestive evidence for the human capital model, and against the migration model. This is also consistent with the results of earlier findings (Barrioluengo and Flisi, 2017; Beine et al., 2020; Soo and Elliott, 2010).

Table 3.4: University-level semi-gravity model to estimate student mobility

	(1) All students	(2) Bachelors	(3) Masters	(4) Doctoral
ln(GDP per capita country)	0.201	-0.145	0.119	1.027*
	(0.446)	(0.342)	(0.491)	(0.587)
Democracy index country	-0.003	-0.003	-0.016	0.055**
	(0.025)	(0.030)	(0.026)	(0.027)
ln(Total core budget)	0.488***	0.357*	0.508***	0.711***
	(0.187)	(0.214)	(0.184)	(0.097)
ln(Student fees funding)	0.376**	0.432**	0.368**	0.241***
	(0.151)	(0.194)	(0.156)	(0.069)
ln(Student to faculty ratio)	0.265*	0.372	0.263**	-0.113
	(0.143)	(0.257)	(0.102)	(0.100)
ln(Ranking score)	0.227***	0.335***	0.176*	0.542***
	(0.080)	(0.098)	(0.098)	(0.194)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Pseudo-R^2	0.73	0.69	0.61	0.82
Observations	934	929	929	866

Note: Standard errors clustered at country level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.6 Survey analysis

3.6.1 Survey details

In order to better understand the forces and motivations that drive students to study abroad, as well as the main factors behind their decision as to whether they return to their home country after graduation or not, I have designed a web-based survey that targets international students.[51]

Before starting the survey design, an extensive literature review was carried out, which identified the factors behind the post-graduation return decision as a gap in the literature. I then designed the first version of the survey such that it comprehensively covers not only this question, but also the motivations behind the student's decision to study abroad. The beta version of the survey

[51]The survey was designed in a broader context, so that it can potentially target international graduates, as well as those who were enrolled abroad but discontinued their studies before graduation. However, since for this chapter the survey has only been run on students, I only focus on the relevant parts of the survey flow here.

was tested in two rounds: first among the faculty of the Department of Economics and Business, second among all doctoral students of the Central European University (CEU). The survey was then fine-tuned according to the feedback received.

The survey is structured as follows.[52] First, it asks about the details of the program the student is enrolled in, as well as details on previous education, if any (university, level, field, tuition fees, scholarships). Next, the survey asks for the importance of several factors that may have affected the students' university choice, as well as whether they have considered home country alternatives as well, and the decisive factors between choosing the foreign and the home university. The next block of the survey concerns the post-graduation plans of the student. The most relevant question in this block is the student's self-assessed probability that they will return to their home country in the next 10 years. The survey also asks for the factors that affect this decision. Finally, the survey ends with demographic questions.

The survey is designed with the long-term aim to have it distributed among the international students of a large number of European universities. As a pilot, the survey was distributed among all CEU students in October 2021, and it was filled in by 600 students;[53] 418 of them have finished the survey completely. 63% of them is studying at the masters level; 29% at the doctoral level; and 8 % at the bachelors level. CEU is a private university that operates campuses both in Budapest, Hungary, and Vienna, Austria. The university defines itself as a diverse and international university. In line with this, the students' background is fairly mixed: their countries of birth covers 84 different countries, with the largest shares accounted for by Hungary (10%), the United States (9%), and Russia (6%).

3.6.2 Survey responses

Regarding the students' motivation for choosing their current university, the survey asked for the importance of several different factors on a scale of 1 to 10. On average, the quality of education ranks the highest (8.8), followed by multicultural environment (8.5) and the reputation of the university (7.8). As regards the quality and reputation of the university, roughly two-thirds have checked the university's position in international rankings prior to applications (primarily QS World University Rankings and Times Higher Education (THE) rankings. These suggest that university-level characteristics are more relevant than the characteristics of the destination country, as the importance of these are ranked considerably lower (7.2 for the importance of living in the country, and 5.4 for the importance of the post-graduation employment opportunities in the country).

Although 67% of the respondents claimed to have not even considered a home country alternative when applying to universities, the survey provides the respondents with a hypothetical scenario of contemplating between the current university and the best home country alternative, and asks for

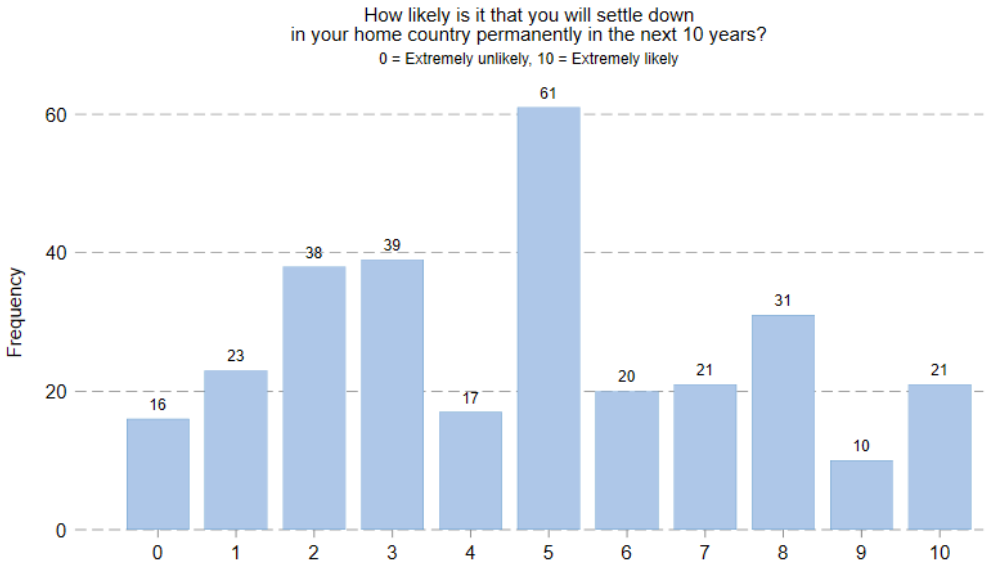
[52] See Annex C.2 for the full survey.

[53] The survey was sent to 2,476 e-mail addresses. 600 responds correspond to a response rate of 24%.

the factors that may have made them choose the home country alternative on a scale of 0 (very unlikely) to 10 (very likely). Consistently with the responses to the questions above, the factors that would make the students choose the home country alternative with the highest likelihood (7.4) is the higher educational quality of the home country universities. This is followed by better financing conditions (6.7), and better political (5.8) and economic (5.7) environment in the home country.

The survey also asks about the likelihood of the student's permanent return to their country of origin in 10 years following graduation. The answers, plotted on Figure 3.2, show a fairly diverse picture, with each option receiving a considerable number of responses. Mapping these answers to a binary category (i.e., those who responded 0-4 are not likely to return, while those responded 6-10 are likely to return; responses of 5 are omitted), the slight majority (56%) of the respondents claim to be unlikely to return their home country in the next 10 years.

Figure 3.2: Distribution of students' return likelihood



Finally, the survey asks for the factors that have the greatest impact on this self-assessed return likelihood. 236 out of the 600 respondents answered this question. Table 3.5 shows the average score of the responses to these factors (on a scale of 1 to 10), broken down into those who are likely and not likely to return (based on the above binary classification). We can see considerable differences between the prospected returners and non-returners, which suggests some clear patterns: for those who are not planning to return, the primary reasons for this are the (lack of) employment prospects (8.46) and the political circumstances (8.38), followed by economic circumstances (8.30). Notably, all these values can be considered very high on a scale of 1-10. For those who are planning

to return, the highest ranked factor behind it is family and personal reasons (8.75).

Table 3.5: Factors affecting return likelihood

Factor	Likely to return	Unlikely to return	Total
Employment prospects	8.14	8.46	8.29
Economic circumstances	7.19	8.30	7.83
Political circumstances	6.75	8.38	7.62
Family or personal reasons	8.75	7.33	7.78
N	*103*	*133*	*236*

To analyse the pattern of return intentions across home countries, I merged the country-level *CEPII*- and *Freedom House*-data on the survey responses, based on the respondents' country of origin. I include three control variables in the model: (i) the natural logarithm of GDP per capita of the home country, (ii) the democracy index of the home country, and (iii) the natural logarithm of the home country's distance from Austria (the primary host country of CEU). All these indicators are included at their 2021 level. Using these control variables, I run four different specifications: the self-assessed return likelihood on a 1-10 scale is estimated both using OLS (Model 1) and Poisson Pseudo-Maximum Likelihood (Model 2); the binary version of the return likelihood is estimated both using linear probability model (Model 3) and logit (Model 4). The results of these estimates are shown in Table 3.6. The results suggest that students are less likely to return to a high-GDP country, and more likely to return to a more democratic country. These findings are not statistically significant, however, which may be attributed to the relatively low sample size. The effect of distance, on the other hand, is highly significant and consistent across specifications: students claim that they are considerably more likely to return to their home country if it is closer to the university.

This result underlines the relevance of the theoretical models to the consequences of student mobility. Individuals that primarily go abroad in order to gain human capital are presumably less likely to take very long distances, as their move is meant to be temporary, and the transaction costs of moving (including non-financials, such as living far from home country social network) do not pay off. Those that (in accordance with the migration model) consider foreign education as a stepping stone to permanent settlement, distance is less of a burden, as it is a one-off investment that is offset by long-term benefits. The survey responses also point to the role that CEU plays in providing opportunities to students from developing countries for permanent settlement in the European Union.

Table 3.6: Regressions estimating return likelihood

	(1) Return OLS	(2) Return PPML	(3) Dummy LPM	(4) Dummy Logit
ln(GDP per capita)	-0.144	-0.032	-0.022	-0.097
	(0.153)	(0.034)	(0.033)	(0.141)
Democracy index	0.008	0.002	0.001	0.003
	(0.007)	(0.001)	(0.001)	(0.006)
ln(distance from Austria)	-0.267**	-0.056**	-0.067***	-0.276***
	(0.123)	(0.025)	(0.024)	(0.105)
R^2	0.02	0.02	0.03	0.02
Observations	296	296	235	235

Note: Heteroskedasticity-robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.7 Conclusions

In this chapter, using data on country- and university-level international student flows, as well as the results of an original, self-designed survey, I investigated the international students' motives for foreign education which, in turn, provides insight to their post-graduation return decision. The results of the country- and university-level analysis suggests that the student flows are more consistent with the predictions of the human capital model than those of the migration model. This predicts a relatively high return rate of the students to their home country, which can be considered as positive from the perspective of the source country. It also has beneficial effect on global convergence: international students returning to their home country can help their country catch up, thus contributing to the reduction of inequalities.

The survey responses by CEU students give a more nuanced picture. A slight majority of the respondents claim that they are not planning to return to their home country in the foreseeable future, mostly for economic and political reasons. If these plans are realized, this is a clear loss for their countries of origin. These survey responses only focus on students of a single university, and as such, these are not suitable to draw generic conclusions regarding the global state of play of international student mobility and return rates. Nevertheless, this chapter, including the designed survey, is an important first step towards the understanding of this topic of high policy relevance.

References

Acemoglu, D., Hassan, T. A., and Tahoun, A. (2018). The power of the street: Evidence from egypt's arab spring. *The Review of Financial Studies*, *31*(1), 1–42.

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., and Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939.

Austen-Smith, D., and Fryer, R. G. (2005). An economic analysis of 'acting white'. *Quarterly Journal of Economics*, *120*(2), 551–583.

Barrioluengo, M. S., and Flisi, S. (2017). *Student mobility in tertiary education: institutional factors and regional attractiveness*. Publications Office of the European Union. https://doi.org/10.2760/675338

Battaglini, M. (2002). Multiple referrals and multidimensional cheap talk. *Econometrica*, *70*(4), 1379–1401.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. In *Human capital: A theorical and empirical analysis*.

Behn, D., Berge, T. L., and Langford, M. (2018). Poor states or poor governance? explaining outcomes in investment treaty arbitration. *Northwestern Journal of International Law and Business*, *38*(3), 1.

Beine, M., Bertoli, S., and Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, *39*(4), 496–512.

Beine, M., Delogu, M., and Ragot, L. (2020). The role of fees in foreign education: evidence from Italy. *Journal Of Economic Geography*, *20*(August 2018), 571–600. https://doi.org/10.1093/jeg/lby044

Beine, M., Noël, R., and Ragot, L. (2014). Determinants of the international mobility of students. *Economics of Education Review*, *41*, 40–54. https://doi.org/10.1016/J.ECONEDUREV.2014.03.003

Bessey, D. (2012). International student migration to Germany. *Empirical Economics*, *42*(1), 345–361. https://doi.org/10.1007/s00181-010-0417-0

Bozóki, A. (1994). Intellectuals and democratization in Hungary. In C. Rootes and H. Davies (Eds.), *Social change and political transformation* (pp. 149–175). Routledge.

Brezis, E., and Soueri, A. (2011). Why do Students Migrate? Where do they Migrate to? *Working Papers, AlmaLaurea Inter-University Consortium*.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., and Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters*, *198*, 109664.

Bursztyn, L., and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, *9*, 131–153.

Campante, F., Durante, R., and Sobbrio, F. (2017). Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association*, *16*(4), 1094–1136.

Carmichael, F., and Thomas, D. (2005). Home-Field Effect and Team Performance: Evidence From English Premiership Football. *Journal of Sports Economics*, *6*(3), 264–281.

Caselli, M., Falco, P., and Mattera, G. (2022). When the stadium goes silent: How crowds affect the performance of discriminated groups. *Journal of Labor Economics*.

Cialdini, R. B., and Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In. G. L. D. T. Gilbert S. T. Fiske (Ed.), *The handbook of social psychology* (pp. 151–192). McGraw-Hill.

Conte, M., Cotterlaz, P., and Mayer, T. (2022). *The cepii gravity database* (Working Papers No. 2022-05). CEPII. http://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=13432

Czernich, N., Falck, O., Kretschmer, T., and Woessmann, L. (2011). Broadband infrastructure and economic growth. *The Economic Journal*, *121*(552), 505–532.

Dányi, E. (2006). Xerox project: Photocopy machines as a metaphor for an "open society". *Information Society*, *22*(2), 111–115. https://doi.org/10.1080/01972240600567212

Dawson, P., and Dobson, S. (2010). The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees. *Journal of Economic Psychology*, *31*(2), 181–191.

DellaVigna, S., and Gentzkow, M. (2010). Persuasion: Empirical Evidence. *Annual Review of Economics*, *2*(1), 643–669.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics*, *118*(3), 909–968.

Dittmar, J. E. (2011). Information technology and economic change: The impact of the printing press. *The Quarterly Journal of Economics*, *126*(3), 1133–1172.

Dohmen, T. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic inquiry*, *46*(3), 411–424.

Dohmen, T., and Sauermann, J. (2016). Referee Bias. *Journal of Economic Surveys*, *30*(4), 679–695.

Duggan, M., and Levitt, S. D. (2002). Winning isn't everything: Corruption in sumo wrestling. *American Economic Review*, *92*(5), 1594–1605.

Enikolopov, R., Makarin, A., and Petrova, M. (2020). Social Media and Protest Participation: Evidence From Russia. *Econometrica*, *88*(4), 1479–1514. https://doi.org/10.3982/ecta14281

Erikstad, M. K., and Johansen, B. T. (2020). Referee bias in professional football: Favoritism toward successful teams in potential penalty situations. *Frontiers in Sports and Active Living*, *2*.

Falck, O., Gold, R., and Heblich, S. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review*, *104*(7), 2238–2265.

Földes, G., and Hubai, L. (1999). *Parlamenti választások Magyarországon 1920-1998*. Napvilág Kiadó.

Garicano, L., Palacios-Huerta, I., and Prendergast, C. (2005). Favoritism under social pressure. *Review of Economics and Statistics*, *87*(2), 208–216.

Gauriot, R., and Page, L. (2019). Fooled by performance randomness: Overrewarding luck. *The Review of Economics and Statistics*, *101*(4), 658–666.

Gentzkow, M., and Kamenica, E. (2016). Competition in Persuasion. *The Review of Economic Studies*, *84*(1), 300–322.

Gong, H. (2022). The effect of the crowd on home bias: Evidence from nba games during the covid-19 pandemic. *Journal of Sports Economics*, *23*(7), 950–975.

Haraszti, M. (2000). Civil kurázsitól civil társadalomig–A magyar szamizdat két évtizede [From civil courage to civil society. Two decades of the Hungarian samizdat]. *Magyar Lettre Internationale*, (38), 55–58.

Hodler, R., and Raschky, P. A. (2014). Regional Favoritism. *The Quarterly Journal of Economics*, *129*(2), 995–1033.

Hodosán, R. (2004). *Szamizdat történetek*. Noran.

Hubai, L. (2001a). *Magyarország 20 xx. századi választási atlasza: 1920-2000*. Napvilág.

Hubai, L. (2001b). *Magyarország XX. századi választási atlasza 1920-2000 I-III.* Napvilág Kiadó.

Jamieson, J. P. (2010). The Home Field Advantage in Athletics: A Meta-Analysis. *Journal of Applied Social Psychology*, *40*(7), 1819–1848.

Kaufman, M. T. (2002). *Soros: The life and times of a messianic billionaire*. Alfred a Knopf Incorporated.

Kaushal, N., and Lanati, M. (2019). *INTERNATIONAL STUDENT MOBILITY: GROWTH AND DISPERSION* (tech. rep.). http://www.nber.org/papers/w25921

Kern, H. L. (2011). Foreign media and protest diffusion in authoritarian regimes: The case of the 1989 east german revolution. *Comparative Political Studies*, *44*(9), 1179–1205. https://doi.org/10.1177/0010414009357189

Lepori, B., Lambrechts, A. A., Wagner-Schuster, D., and Zahradnik, G. (2023). The european tertiary education register, the reference dataset on european higher education institutions. *Scientific data*, *10*(1), 438.

Lutz, W., Amran, G., Bélanger, A., Stonawski, M., Marois, G., Gailey, N., Jensen, K., Ghio, D., Conte, A., Grapsa, E., Muttarak, R., Amran, G., Sabourin, P., and Potančoková, M. (2019). *Demographic scenarios for the EU: Migration, population and education*. EUR 29739 EN, Publications Office, Luxembourg. https://doi.org/10.2760/590301

Manacorda, M., and Tesei, A. (2020). Liberation Technology: Mobile Phones and Political Mobilization in Africa. *Econometrica*, *88*(2), 533–567. https://doi.org/10.3982/ecta14392

Mátyás, L., Bőgel, G., Knell, M., Odor, L., and Weresa, M. A. (2022). Research and development and higher education. In L. Mátyás (Ed.), *Emerging european economies after the pandemic: Stuck in the middle income trap?* (pp. 515–574). Springer International Publishing. https://doi.org/10.1007/978-3-030-93963-2_11

Morita, H., and Araki, S. (2022). Social pressure in football matches: An event study of 'remote matches' in japan. *Applied Economics Letters*, *0*(0), 1–4.

North, D. C. (1994). Economic Performance Through Time. *The American Economic Review*, *84*(3), 359–368.

Nóvé, B. (1999). *Tény/Soros: A Magyar Soros Alapítvány első tíz éve, 1984-1994*. Balassi Kiadó.

Parsons, C. A., Sulaeman, J., Yates, M. C., and Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review*, *101*(4), 1410–1435.

Prendergast, C., and Topel, R. H. (1996). Favoritism in organizations. *Journal of Political Economy*, *104*(5), 958–978.

Racz, B. (1987). Political participation and developed socialism: The hungarian elections of 1985. *Soviet Studies*, *39*(1), 40–62. Retrieved May 26, 2024, from http://www.jstor.org/stable/151435

Ragot, L., Beine, M., and Delogu, M. (2017). *Understanding the Impact of Tuition Fees in Foreign Education: the Case of the UK* (tech. rep.).

Ramos, R. (2024). Modelling migration. In L. Matyas (Ed.), *The econometrics of multi-dimensional panels: Theory and applications* (pp. 455–478). Springer International Publishing. https://doi.org/10.1007/978-3-031-49849-7_14

Reade, J. J., Schreyer, D., and Singleton, C. (2022). Eliminating supportive crowds reduces referee bias. *Economic Inquiry*, *60*(3), 1416–1436.

Romsics, I. C. N. .-. D. . 1. (2003). *Magyarország története a XX. században.*

Rosenzweig, M. R. (2008). Higher Education and International Migration in Asia: Brain Circulation. *Annual World Bank Conference on Development Economics*, 59–100.

Schwartzstein, J., and Sunderam, A. (2021). Using models to persuade. *American Economic Review*, *111*(1), 276–323.

Sebestény, I. (2003). Az egyesületek számbavételének módszertani problémái a történeti statisztikában. *Statisztikai Szemle*, *81*(4), 373–387.

Shorter, L., Brown, S. L., Quinton, S. J., and Hinton, L. (2008). Relationships between body-shape discrepancies with favored celebrities and disordered eating in young women. *Journal of Applied Social Psychology*, *38*(5), 1364–1377.

Silva, J. S., and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, *88*(4), 641–658.

Snyder Jr, J. M., and Strömberg, D. (2010). Press coverage and political accountability. *Journal of political Economy*, *118*(2), 355–408.

Soo, K. T., and Elliott, C. (2010). Does price matter? Overseas students in UK higher education. *Economics of Education Review*, *29*(4), 553–565. https://doi.org/10.1016/j.econedurev.2009.10.016

Soros, G. (1990). *Opening the soviet system*. Weidenfeld; Nicolson.

Steinert-Threlkeld, Z. C., Mocanu, D., Vespignani, A., and Fowler, J. (2015). Online social networks and offline protest. *EPJ Data Science*, *4*(1), 19.

Strömberg, D. (2004). Radio's impact on public spending. *The Quarterly Journal of Economics*, *119*(1), 189–221.

Thissen, L., and Ederveen, S. (2006). *Higher education: Time for coordination on a European level* (tech. rep. No. 68). www.cpb.nl

Tomka, B. (1996). *A magyarországi pénzintézetek rövid története*. Gondolat.

UNESCO. (2023). Unesco institute for statistics (uis).

Van Bouwel, L., and Veugelers, R. (2013). The determinants of student mobility in Europe: the quality dimension. *European Journal of Higher Education*, *3*(2), 172–190. https://doi.org/10.1080/21568235.2013.772345

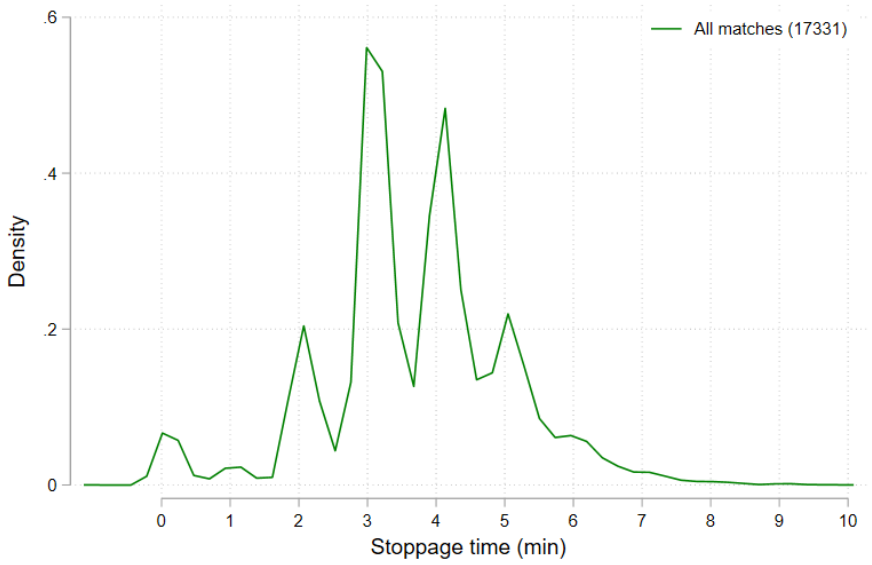
Vollaard, B., and van Ours, J. C. (2022). Bias in expert product reviews. *Journal of Economic Behavior and Organization*, *202*, 105–118.

Wolaver, A., and Magee, C. (2022). Ghost games: Crowds, referee bias, and home advantage in european football leagues. *Journal of Sport Behavior*, *45*(3), 950–975.

Zaslavskaya, O. (2013). Samizdat as social practice and alternative "communication circuit". *Samizdat.*, 87.

DOI: *DOI: 10.14754/CEU.2024.07*

Figure A.1: Distribution of additional time



A Appendix for Chapter 1

A.1 Football rules and practices

A.1.1 Stoppage time

The length of the stoppage time is carefully but not very strictly defined in professional football. According to Point 7.3 of *Laws of the Game*, the official rules of football maintained by the International Football Association Board, the stoppage time at the end of both halves should compensate for the time lost through substitutions, injuries, wasting time, disciplinary sanctions, cooling breaks, VAR checks, and any other cause such as goal celebrations.[54] Thus, the referee decision is driven by a set of detailed rules regarding which events may generate an extension, but the exact length is determined by the referee.[55]

Figure A.1 plots the kernel density estimate of stoppage time distribution across all matches in our sample. We can observe spikes around each minute, the largest at 3 minutes.

[54]The official *Laws of the Game* are available at www.theifab.com/laws-of-the-game-documents/.

[55]See, for example, the following quote by former referee Dermot Gallagher: "*(...)we've had this standardisation that we're going to play 30 seconds per substitution, and for excessive goal celebrations we're to play another 30 seconds – so it starts to tot up, and this is why we find the three or four minutes we have on average at most games.*" – www.playtheadvantage.com/2014/05/27/how-stoppage-time-is-determined/

 DOI: *DOI: 10.14754/CEU.2024.07*

A.1.2 Timestamp of the game end

The event data, collected from whoscored.com, indicated a time stamp for an event type called *end*, at the end of both halves for each game. There may be some delay in the timestamp of this event compared to the actual time of the final whistle, as the system of whoscored.com takes a few seconds to record the final result of the game. To check if this could pose a bias, we took a random subset of 30 games from our analysis sample, games that stood at a one-goal difference before stoppage time across all leagues and seasons. We watched recordings of 30 matches: the game-end timestamp was very close to the actual final whistle, with only minor deviations (ranging from 0 to 6 seconds, with an average of 2 seconds). The example table is available in the paper's code repository.

To see any possible bias, we calculated the average difference between the final whistle and our measured 'end' time across these groups, only to find them the same (2.29 seconds for the home lead and 2.31 seconds for the away lead.).

A.1.3 VAR

The Video Assistant Referee (VAR) system aims to minimize human errors and their influence on match outcomes. Video replays of key events of the game (such as goals, potential penalty situations, and potential red card fouls) are reviewed by an official who communicates with the referee on the pitch. If a potential referee mistake is identified, the game is interrupted for an on-field review of the situation, often lasting several minutes. Time spent reviewing decisions is intended to be compensated for by adding more stoppage time.

A.1.4 Leagues

In a sport season, teams play in their national leagues twice with every other team, once home and once as visitor. A win yields three points, a draw yields one, and a lose yields nothing. These points sum up at the end of season to yield a final league table and create the ranking of teams. As of season 2020/2021, the first four teams from each league will play in the UEFA Champions League, the fifth and sixth will play in the less lucrative UEFA Europa League next season. The last two three teams will be relegated to second divisions. Often a single point decides about winning European spots or relegation.

A.1.5 Big team influence at UEFA

Formally, teams can have no say over referee jobs.[56] However, there were several news reports when teams tried to exert pressure on UEFA regarding referees.[57]

Another area where teams do have informal influence is the format of European games. Indeed, as early as 1997, major teams made the Champions League include runner-ups from 8 leagues, and it was later extended to include the current 4 teams of the top 5 leagues. In 2021, UEFA proposed changes in the Champions League to please the dozen most influential teams that were trying to break away and form the Super League.[58]. The new Champion League format benefits the big teams (mainly English ones).[59]

Football Leaks, a blog, exposed several cases that according to UEFA regulations should have ended with certain penalties but had not in the end, such as for Manchester City and PSG.[60] There are other tax and doping cases of big teams that UEFA helped cover up.[61] These events all suggest a substantial informal influence of big teams on the decision making of UEFA.

A.2 Covid and closures

Due to the Covid-19 outbreak in Europe in spring 2020, practically every football league was suspended as of the second weekend of March. The last round before suspension was played behind closed doors in Italy on the 8th and 9th of March, as well as the last game in our sample, played on the 11th of March in Germany. The remaining games of the 19/20 season were played, with a more intensive schedule, starting as of the 16th of May in Germany, the 11th of June in Spain, the 17th of June in England, and as of the 21st of June in England. Each of these games was played behind closed doors, with no fans allowed to enter the stadium. In France, the remaining games of the 19/20 season were not played.

Season 20/21 started in August 2020 in France and in September 2020 for the rest of the leagues. The vast majority of the games were played behind closed doors. Depending on the severity of the Covid situation, some leagues allowed a restricted number of fans to be present in the stadium for short periods throughout the season. This was the case for France between August and October 2020; for England and Germany between September and October 2020, and in May 2021; for Italy and Spain in May 2021. As Table A.1 shows, this partial opening meant that the stadiums were

[56] See www.documents.uefa.com/r/Regulations-of-the-UEFA-Champions-League-2022/23/Article-48-Appointment-and-replacement-of-referees-Online.

[57] For example, see www.firstpost.com/sports/champions-league-juventus-blames-uefa-chief-refereeing-officer-pierluigi-collina-of-bias-against-serie-a-clubs-4429027.html.

[58] www.mirror.co.uk/sport/football/champions-league-plan-premier-league-26923957

[59] www.forbes.com/sites/steveprice/2022/05/11/uefas-champions-league-changes-benefit-the-big-six-and-newcastle-united/?sh=44446ce57743

[60] www.dailymail.co.uk/sport/sportsnews/article-6347551/Leaked-documents-claim-Manchester-City-hid-30m-UEFA-FFP-investigators.html

[61] www.spiegel.de/international/world/football-leaks-doping-tests-and-real-madrid-a-1240035.html

filled to 10-15% of capacity on average. No match was played with full capacity of fans during season 20/21 – the highest attendance-to capacity ratio in our sample is 33%.

Table A.1: Attendance of matches during Covid

Season	N	N closed	N open	Mean attendance	Max attendance
England 19/20	380	92	0		
England 20/21	380	346	34	12.6	25
France 19/20	279	1	0		
France 20/21	378	316	62	14.3	33.4
Germany 19/20	306	83	0		
Germany 20/21	306	269	37	10.8	20.9
Italy 19/20	380	132	0		
Italy 20/21	380	379	1	1.3	1.3
Spain 19/20	380	111	0		
Spain 20/21	378	373	5	11.8	20.4

Note: The last two columns indicate the mean and maximum, respectively, of attendance-to-capacity percentage among games where reported attendance was greater than zero.

A.3 Data cleaning

For 8 games, the source of our event data contains either no information or obviously erroneous information, such as extremely few events recorded. As we cannot construct the measures of interest for these matches, our analysis excludes them.

Our *Losing offensiveness* measure is not observed for 7 games in our sample, implying that on these games the losing team did not have a single pass in the stoppage time. Thus, we impute a value of 0 for these in our analysis, indicating maximum distance from the opponent's goal line.

A.4 Descriptive statistics

Figure A.2 shows the average stoppage time by goal difference at 90:00. It shows a pattern very similar to the one documented by (Garicano et al., 2005): the average stoppage time is longer for tighter matches. Since goals in stoppage time are rare events, matches with more than one goal difference at the end of the regular time are highly likely to be already settled, and in this case referees tend to blow the final whistle significantly earlier, serving the interest of both teams.

 DOI: *DOI: 10.14754/CEU.2024.07*

Figure A.2: Stoppage time awarded by score margin

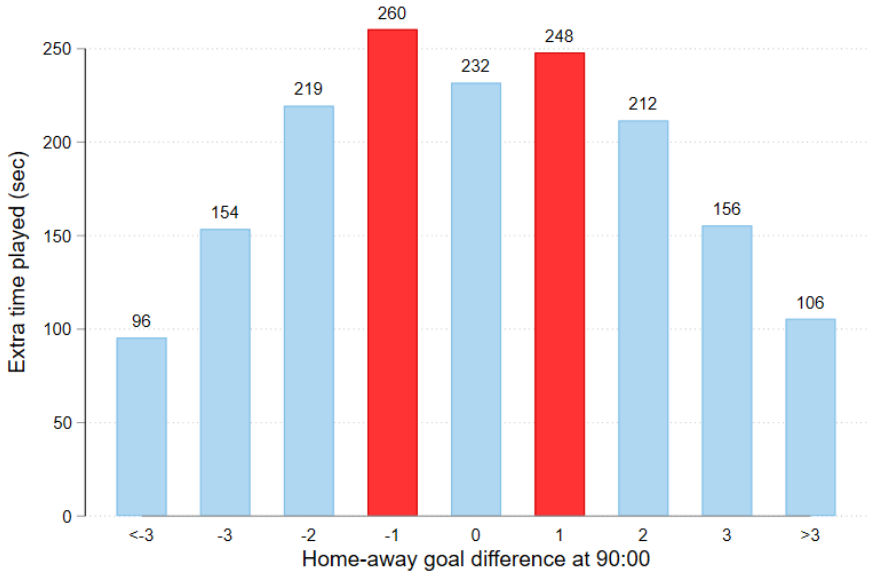


Table A.2 presents some informative descriptive statistics about the variables included in the analysis.

Figure A.3 looks at densities only for games where the home team is winning or losing by a single goal, and we can see that home winning games are characterised by shorter stoppage time than home losing games. A Kolmogorov-Smirnov test also confirms the statistical significance of the difference between the two distributions.

 DOI: *DOI: 10.14754/CEU.2024.07*

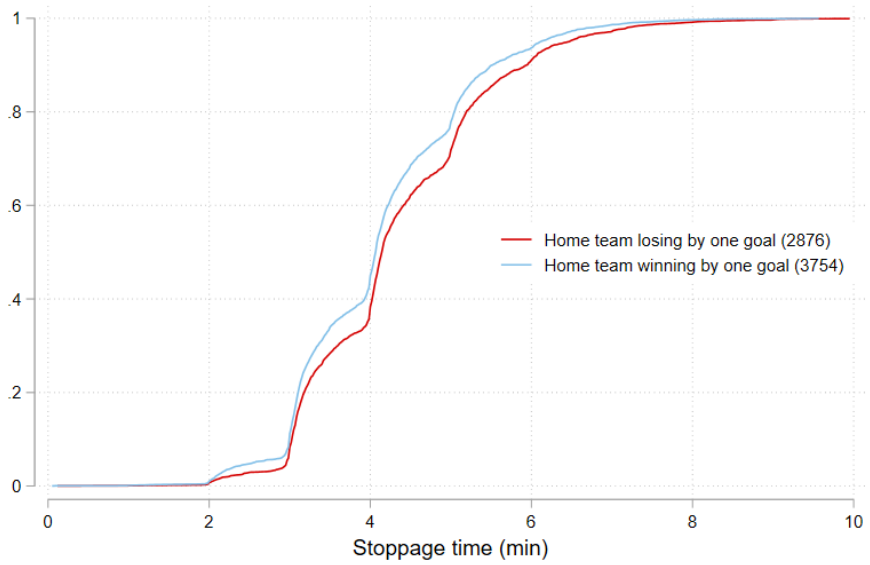
Table A.2: Summary statistics

	Mean	SD	Median	Min	Max	N
Stoppage time (sec)	253.43	69.02	246	3	660	6667
Home losing (0/1)	0.43	0.50	0	0	1	6667
Wasted time (sec)	1244.42	184.55	1241	661	1964	6667
Cards	3.13	1.88	3	0	13	6667
Subs	5.74	1.14	6	2	10	6667
Fouls	14.42	4.18	14	3	33	6667
Goals	1.39	1.13	1	0	7	6667
Losing offensiveness	61.00	9.06	61.56	7.10	100.00	6661
Goals in stoppage time	0.16	0.39	0	0	3	6667
Round	19.17	10.75	19	1	38	6667
Elo ranking difference (home-away)	-8.29	145.12	-5.70	-482.57	477.84	6667
Home rank	10.66	5.55	11.00	1.00	20.00	6667
Away rank	10.23	5.62	10.00	1.00	20.00	6667
Home-away rank difference	0.43	7.93	1	-19	19	6667
Home team value (m EUR)	8.28	9.44	5	1	64	6667
Away team value (m EUR)	9.03	10.21	5	1	64	6667
Value difference (H-A)	-0.75	12.63	-0.17	-61.60	61.60	6667
VAR (0/1)	0.30	0.46	0.00	0.00	1.00	6667

Note: The sample consists of games with a one goal difference at 90:00, from seasons between 2011/12 and 2020/21 of the top 5 European football leagues. See Section 1.2.2 for a detailed description of the variables.

 DOI: *DOI: 10.14754/CEU.2024.07*

Figure A.3: Comparing the cumulative distribution of home team leading vs losing



 DOI: *DOI: 10.14754/CEU.2024.07*

A.5 Additional tables

First, we present the detailed regression results with all control variables.[62] Then, we show a replication of the core table in (Garicano et al., 2005) followed by a table on attendance.

[62]Note that for this table, the parameters we estimate here are not directly meaningful, as the actual number of seconds spent on these events are already partialled out by the wasted time variable. This functional form decision allows for referee heuristics to be accounted for.

 DOI: *DOI: 10.14754/CEU.2024.07*

Table A.3: Detailed regressions indicating home-team bias

	(1)	(2)	(3)	(4)	(5)
Home lose	12.40***	13.38***	11.65***	13.41***	13.40***
	(2.07)	(1.91)	(1.65)	(1.65)	(1.62)
Wasted time		0.18***	0.17***	0.18***	0.18***
		(0.01)	(0.01)	(0.01)	(0.01)
Cards		3.42***	4.75***	4.40***	4.28***
		(0.58)	(0.48)	(0.46)	(0.45)
Subs		4.69***	7.95***	7.39***	7.94***
		(0.85)	(0.84)	(0.81)	(0.78)
Fouls		-3.01***	-2.03***	-2.07***	-1.84***
		(0.28)	(0.21)	(0.21)	(0.21)
Goals		-4.21***	-4.12***	-4.53***	-4.18***
		(0.74)	(0.70)	(0.71)	(0.74)
Losing offensiveness		0.18**	0.25***	0.29***	0.26***
		(0.08)	(0.07)	(0.07)	(0.07)
ET goals		19.13***	18.41***	18.21***	18.04***
		(2.18)	(2.06)	(2.05)	(1.93)
VAR		13.66***	14.46***	14.60***	18.05***
		(2.30)	(1.93)	(1.94)	(2.16)
Round		-0.15**	-0.21***	-0.21***	-0.21***
		(0.06)	(0.06)	(0.06)	(0.06)
Constant	248.04***	13.93			
	(2.48)	(9.21)			
League FE	No	No	Yes	Yes	Yes
Home team FE	No	No	No	Yes	Yes
Referee FE	No	No	No	No	Yes
Observations	6667	6667	6667	6667	6646
R^2	0.01	0.33	0.42	0.45	0.49

Note: Standard errors clustered at home team level.

 DOI: *DOI: 10.14754/CEU.2024.07*

Table A.4: Replication of home bias from Garicano et al. (2005)

	(1)	(2)	(3)	(4)	(5)
Home lose	9.04***	8.94***	8.99***	3.22	11.18*
	(1.69)	(1.75)	(2.97)	(3.34)	(6.20)
Yellow	7.61***	7.71***	7.67***	8.08***	8.06***
	(0.46)	(0.48)	(0.48)	(0.53)	(0.52)
Red	18.38***	18.95***	18.87***		
	(1.75)	(1.67)	(1.73)		
Subs	14.21***	13.92***	14.45***		
	(1.08)	(1.10)	(1.10)		
Home value million	-0.65**	-0.63**	-0.49***	-0.83***	-0.80***
	(0.27)	(0.27)	(0.13)	(0.17)	(0.17)
Away value million	-0.46***	-0.45***	-0.45***	-0.57***	-0.61***
	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)
Home rank	-0.15	-0.19	-0.45*	-0.48*	-0.47*
	(0.32)	(0.30)	(0.25)	(0.28)	(0.28)
Home-away rank diff	0.28*	0.25	0.23	0.34**	0.32*
	(0.16)	(0.16)	(0.16)	(0.17)	(0.17)
Round			-0.16**	-0.21***	-0.22***
			(0.08)	(0.08)	(0.08)
Home lose × Round			0.00		
			(0.12)		
Attendance 1000				234.66***	141.98
				(88.76)	(98.15)
Home lose × Attendance 1000				205.56*	301.01**
				(108.97)	(120.10)
Attendance/Capacity (%)					0.19**
					(0.07)
Home lose × Attendance/Capacity (%)					-0.14
					(0.09)
League FE	Yes	Yes	Yes	Yes	Yes
Referee FE	No	Yes	No	No	No
Season FE	Yes	Yes	Yes	Yes	Yes
Home team FE	Yes	Yes	No	No	No
R^2	0.31	0.36	0.29	0.26	0.26
Model	T2C4	T2C6	T5C4	T6C3	T6C4
Observations	5876	5853	5876	5390	5390

Note: Standard errors clustered at home team level. Controls include time with the ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, whether VAR was used, and the average distance of the passes of the losing team from opponent's goal line in the stoppage time.

 DOI: *DOI: 10.14754/CEU.2024.07*

Table A.5: Explaining attendance/capacity ratio

	(1)	(2)	(3)
Away Top 6	6.09***		
	(0.47)		
Home-away rank diff ≤ -10		-1.68***	
		(0.48)	
Home-away rank diff			0.28***
			(0.04)
Season FE	Yes	Yes	Yes
League FE	Yes	Yes	Yes
Home team FE	Yes	Yes	Yes
R^2	0.76	0.75	0.76
Observations	16075	16075	16075

Note: Standard errors clustered at home team level.

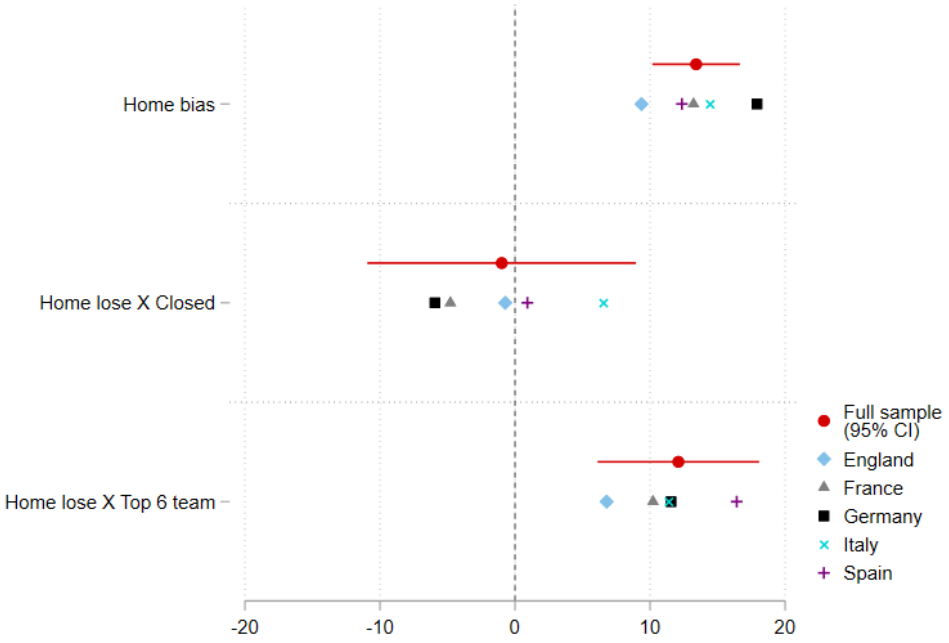
 DOI: *DOI: 10.14754/CEU.2024.07*

A.6 Persistent results across leagues and time

Different countries and leagues may have different customs and regulations. Thus, we might see heterogeneity for any of our results, or find that they are driven by peculiarities in a single country.

To illustrate how heterogeneous our main findings are across countries, we run regressions of Model (4) of Table 1.1 and Model (1) of Table 1.2 for each league separately (without league fixed effects). The estimated coefficients are presented in Tables A.6 and A.7.

Figure A.4: Estimated regression coefficients by league



We find that all main results are highly robust across leagues. First, the home-team bias is very similar, ranging between 9.4 and 17.9 seconds (not statistically different from each other). Second, we see that during closed games, all leagues experienced a small change only, with point estimates ranging between -5.9 and +6.5 seconds, neither being statistically different from zero. Third, in terms of the moderator variable of influential teams, the interaction term of the home lose indicator and the indicator for the top 6 teams is rather stable across leagues, ranging between 6.8-16.4 seconds (neither is statistically different). Note, however, that broken down by leagues, this result lacks statistical power, and the results are not always significant. This is because top teams rarely lose at home, and the number of observations by league is too small. Figure A.4 summarizes our

 DOI: *DOI: 10.14754/CEU.2024.07*

main findings.

Tables A.6 and A.7 show the regressions visualized on Figure A.4.

Table A.6: Regressions by league

	(1)	(2)	(3)	(4)	(5)
Home lose	9.37***	13.22***	17.91***	14.43***	12.36***
	(3.22)	(2.62)	(5.33)	(2.82)	(3.64)
Home lose × Closed	-0.72	-4.79	-5.93	6.54	0.92
	(9.81)	(10.92)	(11.55)	(7.99)	(12.93)
Controls	Yes	Yes	Yes	Yes	Yes
Home team FE	Yes	Yes	Yes	Yes	Yes
Observations	1370	1397	1055	1432	1413
R^2	0.37	0.44	0.43	0.40	0.41
League	England	France	Germany	Italy	Spain

Note: Standard errors clustered at home team level. Controls include time with the ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, whether VAR was used, and the average distance of the passes of the losing team from opponent's goal line in the stoppage time.

Table A.7: Regressions by league

	(1)	(2)	(3)	(4)	(5)
Home lose	7.74*	10.80***	15.12***	12.68***	8.88**
	(4.13)	(3.15)	(4.89)	(3.05)	(4.27)
Home lose × Home Top 6	6.79	10.21*	11.55	11.38*	16.42**
	(6.89)	(5.77)	(7.89)	(5.67)	(6.26)
Controls	Yes	Yes	Yes	Yes	Yes
Home team FE	Yes	Yes	Yes	Yes	Yes
R^2	0.37	0.44	0.43	0.40	0.41
League	England	France	Germany	Italy	Spain
Observations	1370	1397	1055	1432	1413

Note: Standard errors clustered at home team level. Controls include time with the ball out of play, number of cards, substitutions, fouls, goals, goals in stoppage time, whether VAR was used, and the average distance of the passes of the losing team from opponent's goal line in the stoppage time. All these variables are also interacted with the *Closed* dummy.

DOI: *DOI: 10.14754/CEU.2024.07*

B Appendix for Chapter 2

B.1 Timing and magnitude of treatment

This Appendix presents ordinary least squares estimations to capture the difference in treatment effect depending on the timing (that is, the year when the *Xerox* machine arrived in a given municipality) and the magnitude (that is, the number of machines received) of the treatment.

B.1.1 Treatment magnitude

Table B.1 presents the estimation of the following equation:

$$Outcome_i = \alpha + \beta N_i + \gamma X_i + \epsilon_i, \tag{B.1}$$

where N_i stands for the total number of machines given out to settlement i throughout the *Xerox* program.

Table B.1: OLS estimation of the effect of the number of machines

	Yes 1989	Turnout 1990	Postcom 1990	No 2016
Nr of machines	0.116	0.665**	0.362	0.544
	(0.150)	(0.300)	(0.296)	(0.351)
Controls	Yes	Yes	Yes	Yes
Observations	1148	1867	1867	1888
R^2	0.422	0.461	0.326	0.313

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.1.2 Treatment timing

Table B.2 shows the estimation results of the following linear equation:

$$Outcome_i = \alpha + \sum_{k=1985}^{1989} \beta_k T_{ik} + \gamma X_i + \epsilon_i, \tag{B.2}$$

where T_{ik} equals to one if municipality i was provided with machine in year k and X_i is our usual set of control variables.

Table B.2: OLS estimation of the effect of treatment by year

	Yes 1989	Turnout 1990	Postcom 1990	No 2016
Treated in 1985	0.403	1.009	0.356	-0.377
	(0.485)	(1.765)	(0.903)	(1.561)
Treated in 1986	-0.819	1.120	2.978**	2.190
	(0.533)	(1.597)	(1.467)	(1.354)
Treated in 1987	0.194	1.249	0.841	1.408*
	(0.307)	(0.916)	(0.897)	(0.759)
Treated in 1988	-0.072	5.503***	-1.120	5.937***
	(0.486)	(1.556)	(1.822)	(1.658)
Treated in 1989		0.445	2.120**	1.765
		(1.329)	(1.047)	(1.363)
Controls	Yes	Yes	Yes	Yes
Observations	1148	1867	1867	1888
R^2	0.422	0.461	0.328	0.315

Standard errors in parentheses

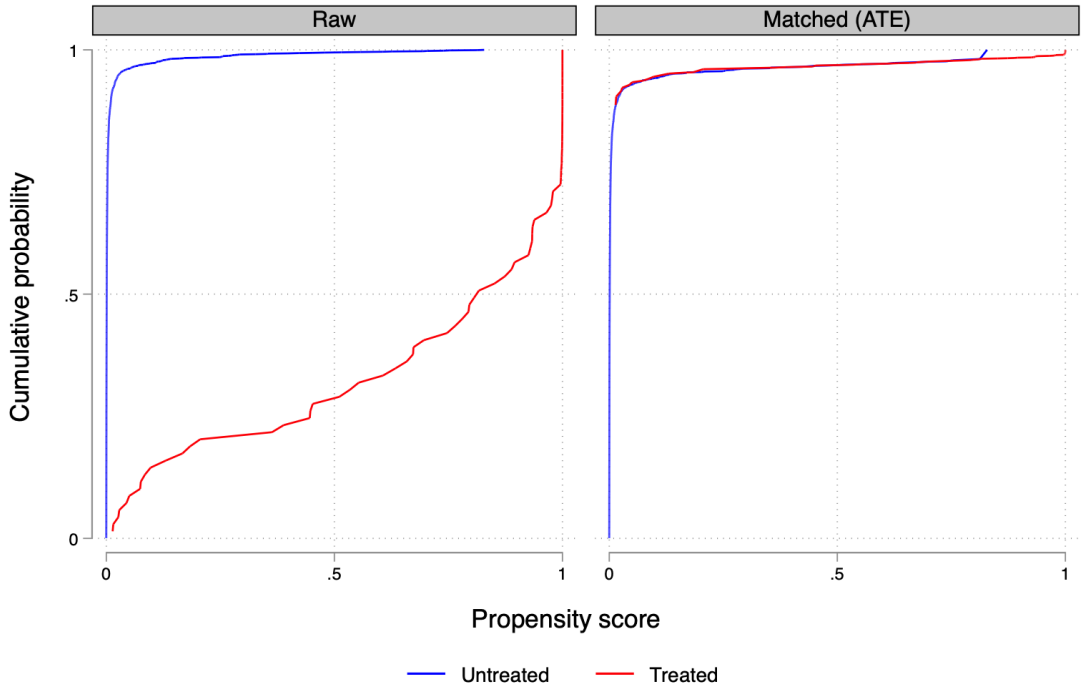
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.2 Balance diagnostics

This section presents some diagnostics related to Section 2.4.2. As documented in Table 2.1, there is considerable imbalance between the control and the treated groups in terms of the pre-determined variables. However, the nearest neighbor propensity score matching does a decent job in balancing the treated and control groups: Figure B.1 plots the cumulative distribution propensity score (estimated with logit) on raw data (left panel) and on the balanced data, that is, after the nearest neighbor matching on propensity score has been carried out (right panel). The common support assumption (also key for the propensity score matching method to work) also holds based on the figure.

 DOI: *DOI: 10.14754/CEU.2024.07*

Figure B.1: Cumulative distribution of propensity score before and after matching



B.3 Robustness check for propensity score matching

We carry out the same analysis as in Section 2.4.2. The main difference is that we keep municipalities without a university, public library, or hospital in the sample. The rationale for this is to check how sensitive our results are to the composition of the untreated control group. As the comparison of Table B.3 and Table 2.1 illustrates, this expansion of the control group does not change the means of the covariates considerably.

 DOI: *DOI: 10.14754/CEU.2024.07*

Table B.3: Balance table on selected covariates. *Note:* Budapest excluded.

Variable	Control	Treated	Diff.
Population (1000), 1880	1.35	13.24	11.90***
Literacy rate (pct.), 1880	41.89	52.44	10.55***
Association members (pct.), 1880	0.62	10.71	10.09***
Capital of financial institutions per capita, 1880	0.82	9.06	8.25***
Turnout 1945	93.66	92.64	-1.02
Communist 1945	12.51	15.58	3.07**
Population (1000), 1980	1.88	41.47	39.59***
Population density (pop/km2)	0.68	4.12	3.45***
% female	52.88	49.97	-2.91***
% active	43.30	49.16	5.86***
% industry	11.30	19.01	7.72***
% agriculture	41.43	12.82	-28.61***
% construction	2.99	3.75	0.76***
% services	11.59	19.74	8.14***
% higher education	1.43	4.93	3.51***
Housing space (m2) p.c.	206.42	83.16	-123.25***
Hospital (y/n)	0.03	0.81	0.78***
Stock of city library, per capita, 1980	2.96	3.74	0.78**
City library subscribers, per capita, 1980	17.82	17.22	-0.60
Observations	3,101	73	3,174

The results of the propensity score matching on the expanded sample are shown in Table B.4. For all outcome variables, the estimated effects are similar to those of the baseline estimates.

Table B.4: Average treatment effects estimated (PSM)

	(1)	(2)	(3)	(4)
	Yes 1989	Turnout 1990	Postcom 1990	No 2016
ATE	2.102***	3.848***	-4.186***	-0.599**
	(0.155)	(0.279)	(0.306)	(0.302)
Observations	1244	2575	2575	2604

Notes: Budapest is excluded.

DOI: *DOI: 10.14754/CEU.2024.07*

C Appendix for Chapter 3

C.1 Robustness check

Table C.1 shows the same regression as 3.4, but without controlling for the ranking of the university, which leads to significantly larger sample size.

Table C.1: University-level semi-gravity model to estimate student mobility (without ranking)

	(1) All students	(2) Bachelors	(3) Masters	(4) Doctoral
ln(GDP per capita country)	0.764*	0.489	1.004**	1.396***
	(0.392)	(0.434)	(0.396)	(0.499)
Democracy index country	0.042	0.031	0.045	0.052**
	(0.034)	(0.036)	(0.034)	(0.023)
ln(Total core budget)	0.676***	0.578***	0.648***	0.925***
	(0.166)	(0.174)	(0.161)	(0.110)
ln(Student fees funding)	0.324*	0.342*	0.341**	0.274***
	(0.166)	(0.193)	(0.163)	(0.087)
ln(Student to faculty ratio)	0.199*	0.357	0.104**	-0.222***
	(0.111)	(0.228)	(0.045)	(0.073)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Pseudo-R^2	0.82	0.74	0.75	0.84
Observations	3589	3441	3350	2272

Note: Standard errors clustered at country level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C.2 The STUMM survey

The questions of the questionnaire are shown below. The questionnaire was distributed on the Qualtrics platform.

 DOI: 10.14754/CEU.2024.07

Start of Block: Introduction

Student Migration and Mobility

Welcome! This survey is the part of a research project which aims to discover the motivations and plans of students who decided to study abroad. We are particularly interested in why you decided to start your higher education studies abroad, what are your plans for the future, and what are the factors that affect these choices.

We plan to circulate a few similar surveys throughout a couple of years in the future, provided that you give your consent for us to do so below.

The survey takes about 15-20 minutes of your time. We do not match your contact information with your responses, so your answers will remain fully anonymous throughout the analysis.

We rely our research project on voluntarily given information, therefore, please, give your consent to the collection and use of your personal data as indicated in the Privacy notice.

Do you consent to these terms?

- ○ Yes
- ○ No

Display This Question:

If We rely our research project on voluntarily given information, therefore, please, give your conse... = No

 DOI: 10.14754/CEU.2024.07

We are sorry that you do not take part in our survey. Unfortunately we cannot record your data unless you consent to the collection of your personal data. Are you sure you DO NOT consent to the collection and use of your personal data?

- ○ Yes
- ○ No

Skip To: End of Survey If We are sorry that you do not take part in our survey. Unfortunately we cannot record your data un... = Yes

Display This Question:

If We are sorry that you do not take part in our survey. Unfortunately we cannot record your data un... = No

We rely our research project on voluntarily given information, therefore, please, give your consent to the collection and use of your personal data as indicated in the Privacy notice.

Do you consent to these terms?

- ○ Yes
- ○ No

Skip To: End of Survey If We rely our research project on voluntarily given information, therefore, please, give your conse... = No

If you consent to further contact, please, indicate the e-mail address where we can reach you within 5 years with the next step of our research project/survey. We will use your e-mail address only for this purpose.

_____

 DOI: 10.14754/CEU.2024.07

Which of the following describes your situation the best?

- ○ I am currently enrolled at a university, not in my native country
- ○ I have graduated from a university not in my native country
- ○ I was enrolled at a university not in my native country, but I discontinued my studies before graduation

End of Block: Introduction

Start of Block: ST - STUDIES

In which country do you currently reside?
Country:

▼ Afghanistan ... Zimbabwe

At which university are you currently enrolled? Start typing the English name or the city of the university and choose from the list. (Type the name of the university if not on the list.)

In which year did you enrol in your current university?
Year

▼ 2000 ... 2021

What is your field of study?
Field

▼ Anthropology ... Other

 DOI: 10.14754/CEU.2024.07

What is the primary language of your program?

- ○ English
- ○ Other: _____

What is the level of your program?

- ○ BA/BSc
- ○ MA/MSc
- ○ PhD

Display This Question:
If What is the level of your program? = BA/BSc

Is your current university enrollment your first one?

- ○ Yes
- ○ No

In which year are you expected to graduate?
Year

▼ 2021 ... 2030

Display This Question:
If What is the level of your program? = MA/MSc
Or What is the level of your program? = PhD

 DOI: 10.14754/CEU.2024.07

In which country and year did you complete your BA/BSc studies?
Country
Year

▼ Afghanistan ... Zimbabwe ~ 2022

Display This Question:
If What is the level of your program? = PhD

In which country and year did you complete your MA/MSc studies?
Country
Year

▼ Afghanistan ... Zimbabwe ~ 2022

Do you pay any tuition fee for your current studies?

- Yes
- No

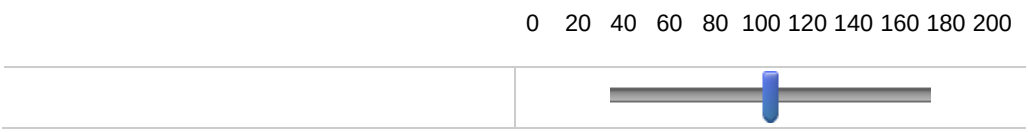
Do you receive any scholarship for your current studies?

- Yes
- No

Display This Question:
If Do you pay any tuition fee for your current studies? = Yes
And Do you receive any scholarship for your current studies? = Yes

Approximately what percentage of the tuition fee does your scholarship cover?
Percentage

 DOI: 10.14754/CEU.2024.07



Do you work besides your studies?

- ○ No
- ○ Yes, part-time
- ○ Yes, full time
- ○ Yes, casually

Display This Question:
If Do you work besides your studies? = Yes, part-time
Or Do you work besides your studies? = Yes, casually

*

On average, approximately how many hours do you usually work in a month?

_____

End of Block: ST - STUDIES

Start of Block: ST - WHY

In which country were you born?
Country

▼ Afghanistan ... Zimbabwe

 DOI: 10.14754/CEU.2024.07

In which country did you complete your secondary education?
Country

▼ Afghanistan ... Zimbabwe

When did you leave your country of origin?
Year

▼ 1990 ... 2021

Did you move with your family (parents, spouse, etc.)?

- Yes
- No

Display This Question:
If Did you move with your family (parents, spouse, etc.)? = Yes

Did your educational prospects play a role in your family's decision?

- Yes
- No
- I don't know

How important were the following factors when you applied to your current university?

	Not at all									Extremely
	0	1	2	3	4	5	6	7	8	9	10

 DOI: 10.14754/CEU.2024.07

Educational quality	
Reputation of the university	
Employment opportunities upon graduation	
Living in that particular country	
Language of instruction	
Multicultural environment	
Friends living near university	

Was your current university your first choice?

- ○ Yes
- ○ No

Display This Question:

If Was your current university your first choice? = No

Which university was your first choice? Start typing the English name or the city of the university and choose from the list. (Type the name of the university if not on the list.)

Did you check the position of your current university in international rankings before you applied?

- ○ Yes
- ○ No

 DOI: 10.14754/CEU.2024.07

Display This Question:

If Did you check the position of your current university in international rankings before you applied? = Yes

Which ranking(s) did you check?

- ☐ Times Higher Education (THE)
- ☐ QS World University Rankings
- ☐ Academic Ranking of World Universities (ARWU) / Shanghai Ranking
- ☐ US News and World Report
- ☐ Other(s): _____

Display This Question:

If Did you check the position of your current university in international rankings before you applied? = Yes

What role did these rankings play in your choice?

Not important at all Extremely important

0 1 2 3 4 5 6 7 8 9 10

Did you also consider universities at your home country?

- ○ Yes
- ○ No

 DOI: 10.14754/CEU.2024.07

Display This Question:

If Did you also consider universities at your home country? = Yes

Which home country university would have you most likely applied/enrolled? Start typing the English name or the city of the university and choose from the list. (Type the name of the university if not on the list.)

Display This Question:

If Did you also consider universities at your home country? = Yes

Where do you think your home country alternative stands in terms of academic ranking compared to your current institution?

- ○ Much lower
- ○ Slightly lower
- ○ About the same
- ○ Slightly higher
- ○ Much higher

Display This Question:

If Did you also consider universities at your home country? = Yes

In terms of financial circumstances (tuition fees / scholarship), how your home country alternative would have compared to your current institution?

- ○ Much more favorable
- ○ Somewhat more favorable
- ○ About the same
- ○ Somewhat worse
- ○ Much worse

 DOI: 10.14754/CEU.2024.07

Display This Question:

If Did you also consider universities at your home country? = Yes

Imagine your younger self, contemplating between your current university and your best home country alternative.
What do you think, to what extent could the factors below have made your younger self choose your home country alternative, rather than your current university?

	Very unlikely 0 1 2 3 4 5 6 7 8 9 10 Very likely
better financial (tuition fee / scholarship) conditions	
higher educational quality	
better economic environment in home country	
better political environment in home country	
other:	

Did you also consider universities in other foreign country?

- ○ Yes
- ○ No

Display This Question:

If Did you also consider universities in other foreign country? = Yes

In which country would have been your most preferred alternative university?
Country

▼ Afghanistan ... Zimbabwe

 DOI: 10.14754/CEU.2024.07

Display This Question:

If Did you also consider universities in other foreign country? = Yes

Why did you end up not enrolling there? (Select all that apply.)

- ☐ did not get accepted
- ☐ worse financial (tuition fee / scholarship) circumstances
- ☐ poorer educational quality of university
- ☐ worse economic circumstances of country
- ☐ worse political circumstances of country
- ☐ other: _____

End of Block: ST - WHY

Start of Block: ST - COETHNIC

Did you contact anybody living in your current city / studying at your current university when you were considering applying?

- ☒ No
- ☐ Yes, people of my nationality
- ☐ Yes, people of not my nationality

Do you have any friends that you moved together with?

- ○ Yes
- ○ No

 DOI: 10.14754/CEU.2024.07

Display This Question:
If Do you have any friends that you moved together with? = Yes

Was this an important factor in your decision to study abroad?

- Yes
- No
- I don't know

Do you socialize with people of your nationality?

- Often
- Sometimes
- Never

Display This Question:
If Do you socialize with people of your nationality? = Often
Or Do you socialize with people of your nationality? = Sometimes

Did you mostly know these people before you started your studies?

- Yes, I knew them before
- No, we first met during my studies

End of Block: ST - COETHNIC

Start of Block: ST - PLANS

 DOI: 10.14754/CEU.2024.07

What are your plans after graduation?

- ○ Study further
- ○ Start working
- ○ I don't know yet

Display This Question:
If What are your plans after graduation? = Study further

In which country are you planning to study further?

- ○ In my home country
- ○ In the country I am currently living
- ○ I don't know yet
- ○ Other: _____

Display This Question:
If What are your plans after graduation? = Study further

In which university are you planning to study further? Start typing the English name or the city of the university and choose from the list. (Type the name of the university if not on the list.) Leave it empty if you do not know yet.

Display This Question:
If What are your plans after graduation? = Start working

 DOI: 10.14754/CEU.2024.07

In which country do you plan your first employment after graduation to take place?

- ○ In my home country
- ○ In the country I am currently living
- ○ I don't know yet
- ○ Other: _____

Display This Question:
If What are your plans after graduation? != Start working
And What are your plans after graduation? != Study further

In which country do you plan to live in the first year after graduation?

- ○ In my home country
- ○ In the country I am currently living
- ○ I don't know yet
- ○ Other: _____

How likely is it that you will settle down in your home country permanently in the next 10 years?

Extremely unlikely Extremely likely

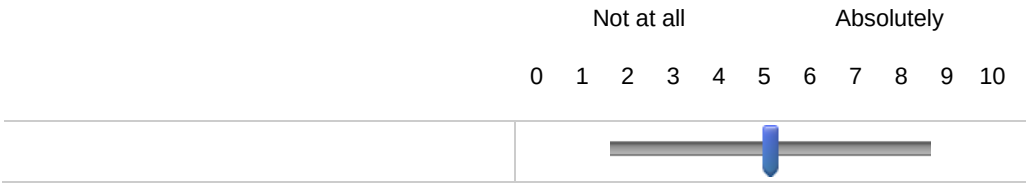
0 1 2 3 4 5 6 7 8 9 10

How important are the following factors when you consider returning to your home country?

Not at all Extremely

DOI: 10.14754/CEU.2024.07

How certain are your above plans?



*

What is your approximate expected monthly net starting salary (in EUR)?

_____

End of Block: ST - PLANS

Start of Block: DEMO

 DOI: 10.14754/CEU.2024.07

Overall, how satisfied are you with your decision to move abroad?

1
2
3
4
5

What is your gender?

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ Prefer not to say

When were you born?
Year

▼ 1970 ... 2021

End of Block: DEMO

Start of Block: FEEDBACK

 DOI: 10.14754/CEU.2024.07

Thank you for your cooperation! At the end of the survey, let us ask how would you grade (from A+ as best to F as worst) the overall quality of this questionnaire.

1
2
3
4
5
6
7
8
9
10
11
12
13

We would appreciate any kind of feedback and suggestions in the box below.

_____

_____

_____

_____

_____

End of Block: FEEDBACK