
# Essays on Urban and Development Economics

Louise Alice Karine Bernard

Department of Geography and Environment

London School of Economics and Political Science

A thesis submitted to the Department of Geography and Environment of the London School of Economics for the Degree of Doctor of Philosophy

London, November 2022

## Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).



I declare that my thesis consists of approximately 40,000 words, excluding all appendices and the bibliography.

## Statement of co-authored work

I certify that Chapter 2 of this thesis is co-authored with Allan Beltran and Karlygash Kuralbayeva, and I contributed 75% of the work. Chapter 3 of this thesis is co-authored with Capucine Riom, Paolo Avner, Jun Rentschler and Rui Su and is part of a larger research project. I contributed 40% to the writing of this output in its current form.

## Acknowledgements

Doing a PhD is hard. It is precisely because it is not an easy path that I would like to thank the many people that have made this PhD possible. I am thankful for the financial support of the ESRC and the London School of Economics.

I would like to thank my supervisors, Henry Overman and Vernon Henderson, for providing guidance and supervision during my PhD.

This thesis has greatly benefited from interactions in and out of weekly seminars, conferences and workshops. I am grateful to participants of the Urban Economics Association (UAE) annual meetings and PhD Summer school, the Association of Environmental and Resource Economists (AERE) meeting, the University of Birmingham and King's College London economics department work-in-progress seminars. In the department of Geography at LSE, I am particularly indebted to Gabriel Ahlfeldt, Felipe Carrozi, Steve Gibbons, Christian Hilber, Lindsey Rehilan and Olmo Silva for their helpful comments during work-in-progress seminars and office hours.

This thesis is also the product of many hours of discussion and work with my co-authors: Paolo Avner, Allan Beltran, Karlygash Kuralbayeva, Jun Rentschler, Capucine Riom and Rui Su. A special thanks to Capucine for her support during the last days of the PhD submission process.

Among all, I am so grateful for the interactions with my peers. It all started with Dzahmilya and Tanner, with whom I spent long days learning the wonders of GIS processing and economics. I also remember the EC400 times with Margarida. I obviously could not forget my fellow climbers Andreas, Giorgia, Lindsey as well as LSE residences colleagues George, Lars, Matt. Lunch and Friday drinks would not have been the same without Antonio, Julien, Glen, Sanchayan, Sandro, Vivian, Yohan and many others.

Writing the PhD was a challenging experience, especially during the COVID-19 pandemic. However, I always could count on my friend, fellow PhD student, climber, cook and amazing woman all around Tiernan to cheer me up. And, of course, I will never forget my two lockdown supporters, Becky and Aaron, whose countless challenges and workouts made the difference.

Lastly, I am grateful to my parents and family for giving me the intellectual curiosity and strength to do this programme.

## Abstract

This thesis is composed of four essays on urban and spatial economics. The first two papers are empirical studies evaluating the impact of public policies in England – one looking at transport infrastructure and the other at flood management. The last two papers leverage satellite imagery to investigate the effects of floods and flood risk on urbanisation in developing countries.

The first paper focuses on the impact of cycling infrastructure on road traffic in London. It demonstrates that providing segregated cycling lanes increases cycling flows without impacting motorised traffic. Not only do the cycling flows increase immediately after the opening of the dedicated lanes, but they also appear to be on a permanent steeper growth path. One primary causal mechanism investigated is the reduction in accidents along the cycling routes.

The second paper analyses the role of natural disasters in local election results in England. It finds that at the electoral ward level, electors punish the incumbent party after a flood during local elections in England – but they are much more likely to do so if the incumbent party aligns with the party in power, both at the local authority and national government levels. There is no evidence that the political party alignment of the incumbent is a significant driving force. However, there is a clear pattern of more votes going to the UK Independence Party in the wake of a flood shock.

The third paper of the thesis investigates the causal role of land scarcity and path dependence on the expansion of Chinese cities into high flood risk land. It finds that a naïve OLS regression overestimates the role topographic constraints play in driving urbanisation in high flood risk areas. Once instrumented for, land scarcity due to topographic constraints is not a driver of urbanisation in high flood risk areas: cities expand into high flood risk land despite having safe land to expand on.

The last paper explores the medium-term effect of flooding on population growth in Sub-Saharan Africa. It finds that large floods in rural areas have long-term persistent effects on population growth but that the effects are mitigated in large urban areas. Using Demographic and Health Survey data, the paper finds that experiencing a severe flood is associated with worse health outcomes and a higher probability of being classified in the poorest wealth bracket, especially in rural areas. In the medium-term, the analysis shows sorting of the poorest households in high-flood risk areas. This is consistent with a higher out-migration rate from rural areas.

# Contents

# List of Figures

# List of Tables

# 1 Introduction

This thesis is composed of four essays on urban, development and spatial economics. The first two papers are empirical studies evaluating the impact of public policies in England – one looking at transport infrastructure and the other at flood management reform. The last two papers leverage satellite imagery to investigate the effects of floods and flood risk on urbanisation in developing countries.

This first paper investigates the impact of constructing segregated cycling lanes on cycling and motorised traffic. I use an event study analysis on the progressive roll-out of the Cycle Superhighways (CSH) programme in central London from 2014 to 2019. Analysing the impact of lane segregation on traffic flows is challenging as cycling lanes location is an endogenous choice from the policymakers and stakeholders. Lanes have been built in strategic locations due to user demand, wide enough roads to accommodate the additional segregated lanes, safety and potential for growth. However, even after considering the endogeneity of the locations using an event study type of analysis, I find that the segregated lanes programme increased cycling traffic flow by about 25% after opening and then by an additional 20% per year. It accounts for nearly half of the increase in cycling traffic along those roads. There is little evidence that this effect is driven by the displacement of cyclists from nearby roads, which indicates that the new lanes increase the number of cyclists or their frequency. One of the mechanisms investigated is an increase in cycling flow via safer trips and safety in numbers. I find a substantial decrease in accidents involving cyclists after the lanes' construction. I do not find any impact on car traffic.

The second paper shows the relationship between natural disasters, decreased support for government parties, and increased votes for protest parties. Failure to prevent floods has considerable physical and economic costs. It also has long-term effects on the well-being of impacted people. We analyse how voters respond to such disasters in the context of "second-order" elections – elections with lower turnout and media coverage. Considering the characteristics of local elections in the UK – half the turnout of General Elections, a higher share of votes for non-dominant parties, and poor knowledge of constituents of local councils

functioning, we show evidence that voters use floods to punish parties in power and give rise to protest parties. We find that electors punish the incumbent party after a flood during local elections in England – and they are much more likely to do so if the incumbent party aligns with the party in power, both at the local council and the national government level. We also find that recent floods bolster the rise of the UK Independence Party. While the baseline decrease in the share of votes for the incumbent is relatively small – 0.7 percentage point, it can be increased by up to 6 percentage points if the incumbent is from the same party as the national government. These effects are more substantial for larger floods in terms of the flooded area and duration. We also find the same bounded rationality as previous papers in the literature: the effect of floods on local electoral outcomes is limited to floods within a year of the election.

In the third paper, we overlay new high-resolution satellite data and flood hazard maps to study how cities expanded in high flood risk areas in China from 1985 to 2015. We construct a panel of cities to investigate the causal role of land scarcity and path dependence on the expansion of cities into high flood risk land. Land scarcity around a city is partly an endogenous product of urbanisation, as cities expand in the direction of unviable land (water bodies or steep land) for their amenity or productivity value, and partly an exogenous process due to the location and size of topographic constraints as cities expand. We develop an instrument for land scarcity based on topographic obstacles encountered by mechanically growing cities, building on Harari (2020). We find that a naïve OLS regression overestimates topographic constraints' role in driving urbanisation in high flood risk areas. Once we use long-difference to remove time-invariant effects and instrument for endogenous growth, we find that land scarcity due to topographic constraints is not a driver of urbanisation in high flood risk areas: cities expand into high flood risk land despite having safe land to expand on. We document how these drivers vary by flood risk types (pluvial, fluvial and coastal).

The last paper explores the medium-term effect of flooding on population growth in Sub-Saharan Africa. To circumvent the lack of comprehensive reporting of floods in low-income countries, I use a flood model and climate data to create a novel historical dataset on floods. I then construct a panel dataset of population, degree of urbanisation and floods from 1990

to 2015 in Sub-Saharan Africa. The satellite imagery allows for the analysis of the effect at a high degree of disaggregation. I find that large floods in rural areas have long-term persistent effects on population growth, but the effects are mitigated in large urban areas. Using Demographic and Health Surveys, I explore the mechanisms behind these diverging trends. In the short-term, rural households are much more likely to suffer negative health and wealth impacts from floods. In the medium-term, sorting and out-migration of poorer households in risky areas are much stronger in rural areas.

# 2 The impact of segregated cycling lanes on road users

## 2.1 Introduction

The world's major cities have built cycling infrastructures in the last decades. Active travel, including walking and cycling, has been encouraged to reduce motorised traffic, bring health benefits and reduce air pollution. In addition, these modes can provide relief to congestion of public transport in central areas. However, these benefits are conditional on doing more than merely displacing bike users from one lane to another, as well as generating a genuine shift in modal share and not only capturing the population's growth in these areas.

Although major cities have spent considerable amounts on increasing the number of cyclists, segregated cycling lanes have not been studied by economists as extensively as other infrastructures such as segregated bus lanes. London's cycling policy reflects other large metropolitan areas in implementing bike-sharing systems, segregated lanes and encouraging users to avoid car traffic. However, bikes represent still only 2.5% of trips in London. London Mayor's strategy is to reach 5% by 2026 (Transport for London 2018). The current strategy aims at convincing more people to cycle. Safety concerns are the first deterrent for cycling in London. Segregated cycling lanes aim directly at improving safety by separating traffic from cars and providing safe junctions. Their design in London - large straight roads from outer neighbourhoods to the centre of the city - was chosen to be easily recognizable and offer a fast and simple way to travel across central London.

This paper examines whether building cycling infrastructure increases cycling flows in large cities. I use the Cycle Superhighways (CSHs) programme in London to conduct the analysis: twelve planned segregated cycling lanes that commuters use to commute safely from the outer neighbourhoods of London to the centre. The main difficulty in studying this programme is that the lane placement was chosen to maximize existing cycling flows. It is thus difficult to disentangle existing trends from the impact of the cycling infrastructure. The program's roll-out between 2008 and 2020 is used to address identification issues. In 2008, Mayor Ken Livingstone announced London's Cycle Superhighways scheme (CSH) as shown in Figure 2.A.1.

By 2010, the first lanes were built[1], but they were perceived as unsafe by users. In response to the criticism, in 2012, the first segregated cycle lanes were built[2]. Most of the subsequent lanes were completely separated from the car traffic[3]. The analysis focuses on the second generation of segregated cycling lanes for practical reasons (I do not observe pre-trends in the first generation of lanes) and to narrow the analysis to mainly segregated lanes.

The paper finds evidence of a net increase in cycling flows for three years following the launch of the program. There is not enough data to estimate longer time effects. As soon as the facility opens, cycling flows increase by about 25%, and then by 20% per year after that. From the empirical design, I find that the increase is not driven by population growth in these areas. I also pay particular attention to the possibility of cycling lanes displacing other traffic. Indeed, the increase in cyclist flow on CSHs could be driven by cyclists choosing safer lanes to do their usual trips. While not a bad outcome in itself, the main goal of the CSH was to create incentives for people to cycle and do more trips cycling. I find no evidence of cycling displacement around the new segregated cycling lanes. The lack of displacement indicates that the increase in traffic in segregated lanes is likely to contribute significantly to the net increase of cycling flow in London, rather than shifting existing flows from other routes. It makes sense as cycling trips are generally short[4], and any additional detour would significantly decrease the advantage of using a bike.

I reproduce the analysis using the London cycle hires - London's public bicycle hire scheme opened under Boris Johnson's mayorship - and show a similar pattern. Following the opening, trips starting or ending at CSHs show an increase while stations further away do not.

Another form of displacement could be car traffic. The reduced effects on pollution would be voided. To disentangle these effects, I analyse the impact of the segregated cycling lanes on cycling at different distances of the lanes. I do not find evidence of decreased car flow or bus flow in the lanes that have been reduced to accommodate the segregated cycling lanes nor in

[1] CS3, CS7 in 2010 and CS2, CS8 in 2011

[2] CS5, upgrade of CS2, the extension of CS3, CS6, CS1 + "Better junctions"

[3] I show the differences between the two generations of lanes in Figure 2.A.2 and 2.A.3. The lane number, e.g. CSH8/CSH5, corresponds to the original plan number and not the order of construction; some lanes are called "CS" and others "CSH"

[4] 22 min on average in London (LTDS, 2018)

the adjacent lanes.

Finally, I look at the underlying mechanism for the increase in cycling, such as the increased safety of cyclists in segregated lanes. I find that these infrastructures bring direct benefits by decreasing the number of car-cycle accidents and reducing accidents per cyclist.

The main contribution of this paper is to capture that segregated cycling lanes do not offer only a one-time increase in cycling at opening but also put cycling flow on a higher growth path. I further demonstrate that the increase is not due to the displacement of cyclists. I also show that the main argument against segregated cycling lanes - disruption of car traffic - did not manifest in London. Finally, I find a significant decrease in accidents after the opening of lanes - explaining the success of the programme.

This paper has clear policy implications. First, the findings imply that building cycling infrastructure has an immediate impact on traffic flows and that this impact is growing over time. The impact is more prominent for fully segregated lanes but still considerable for partially non-segregated lanes. A cost-benefit analysis should take into a large time frame to evaluate these programmes. Additionally, cycling lanes are also often criticised for increasing congestion. However, in the analysis, I find little evidence of change in traffic flows around the cycling lanes. Finally, in line with the previous literature, this paper provides evidence of safety in numbers for cyclists. Not only does the number of accidents per cyclist on the road decrease, but the number of total accidents also drops after the construction of the cycling lanes.

The rest of the paper is structured as follows. First, I review the literature on transport in London and cycling in Section 2.2. I then present the empirical analysis in Section 2.3 and the datasets I use in the analysis in Section 2.4. In Section 2.5, I decompose the results between cycling flows on the new segregated lanes, the displacement analysis on neighbouring roads and the traffic accidents analysis. Finally, I summarise the results and alleys for future research in Section 2.6.

## 2.2 Literature review

Transport economists are no strangers to London's setting. London is famous for experimenting with a central congestion charge in the early 2000s. The effects of the congestion charge were wide-ranging: reduced motorised traffic, decreased air pollution and accidents, increased housing prices and increased traffic outside the congestion zone. Leape (2006) summarises the early implementation of the London congestion charge reduced all motorised traffic by 12% and up to 34% for cars. Green, Heywood, and Navarro (2018) show evidence that London's Congestion Charge reduced traffic accidents and air pollution in the tolled zone. Keat Tang et al. (2016) uses a partial equilibrium to find an elasticity of housing values with respect to traffic of -0.3. More recently, Herzog (2020) uses a general equilibrium model to show that the congestion charge reduces both the number of commuters and their propensity to drive inside the congestion zone but increases driving among untolled drivers. This paper contributes to the economic literature by evaluating the impact of new transport infrastructure on transport mode and its general impact on motorised traffic.

On segregated cycling lanes, studies have shown they are safer for cyclists (Cohen 2013; Li, Graham, and Liu 2017; Mulvaney et al. 2015; Reynolds et al. 2009; Aldred et al. 2018). These studies highlight a few caveats that are worth noting. First, there is a learning period when new lanes are introduced as users learn how to use them safely. Second, there is safety in numbers, meaning that cycling infrastructure might be particularly useful to sustain a higher cycling growth rate. Third, safer infrastructure is also more inclusive: women, young people or the elderly are more likely to cycle when cycling routes are separated from car traffic. Finally, these studies do not consider that there could be endogeneity in cycling lanes placement and existing pre-trends. Therefore, it is essential to show that these results hold even when pre-existing trends are considered.

In London specifically, Aldred et al. (2017) review the literature on cycling provisions separated by motor traffic. They find that even though all users prefer separation, women have stronger preferences. In a follow-up paper, Aldred and Dales (2017) show that the lack of infrastructure in London and the high-level of perceived danger is a deterrent for most casual users. A study by Li et al. (2018) relates an increase in cycle hire near the CSHs. I

generalise the analysis to cyclists using their own bikes, as cycle hire users differ from the general population - cycle hires are used more by tourists and casual users. Li, Graham, and Liu (2017) find no impact of the programme on traffic accidents in the first phase of the programme (2007-2014). They note that a significant drawback of the CSH programme is the lack of separation between cars and cyclists. Bhuyan et al. (2021) uses propensity score to evaluate the impact of CSHs on traffic congestion and find a positive impact of the programme. This paper focuses on the programme's second phase when most routes were built with a physical separation from the motorised traffic in reaction to the early criticism. I find a substantial impact on the reduction of traffic accidents per cyclist.

### 2.3 Estimation strategy

In this paper, I study the impact of the construction of the CSHs on cycle traffic in London using an event study analysis. The treatment group is sites with an active CSH; the control group is not yet treated sites and treated late (sites that opened in 2020).

The set-up behind this paper is that individuals in London have a large set of options regarding modal choice. They can choose to walk, cycle, take public transport, hire taxis or private cars. The determinants of modal choice depend on the individuals, trips and the modes' characteristics. Intuitively, building segregated cycling lanes reduces the cost of travelling by bike. It might also increase the cost of travelling by car by reducing road capacity. In consequence, it should increase the demand for cycling and potentially create a substitution with other modes.

One way to investigate this increase in demand would be to use travel diaries. They exist for London (London Travel Demand Surveys from Transport for London), but unfortunately, the level of geographic disclosure is too aggregated to perform this analysis. In this paper, I thus present results on cycling demand increase, but I cannot comment on substitution or general equilibrium effects.

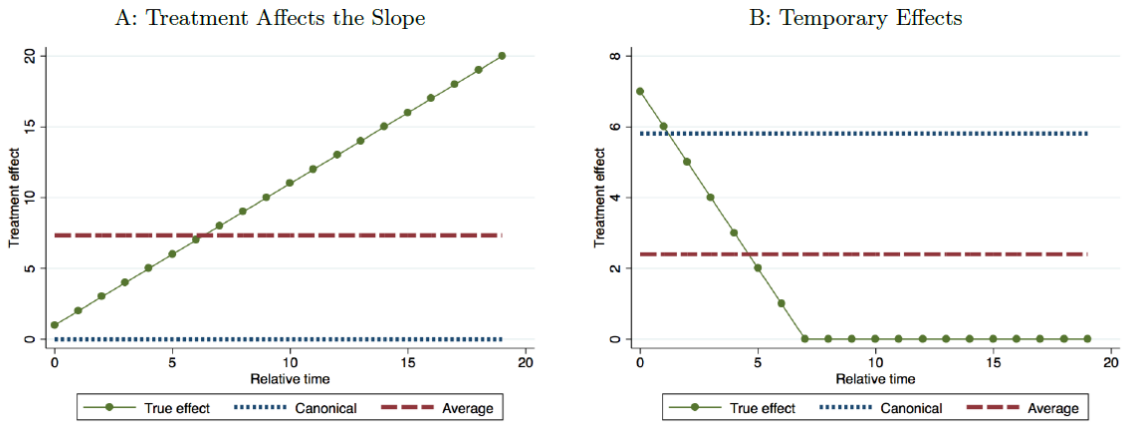
To conduct the analysis, I use the cycle monitoring programme created by Transport for London. It contains various yearly and quarterly surveys available from 2014 to 2019 to track

cycling volume across Central London over time. To measure the impact on motorised traffic, I use a similar geocoded survey produced by the Department for Transport road counts for Greater London from 2000 to 2019. I also use the STATS19 dataset from the Department of Transport that records all road accidents with police involvement in the past decades. These datasets contain the exact geo-coordinates of the counting sites and accidents that I spatially relate to the CSHs routes.

The programme's specific design allows for overcoming some of the issues highlighted by the recent literature on the difference in differences (DiD) in staggered adoption (Sun and Abraham 2018; Borusyak and Jaravel 2018; De Chaisemartin and D'Haultfoeuille 2018; Goodman-Bacon and Marcus 2020). The canonical difference in differences estimator has two time periods: before and after implementation, and two groups: treatment and control. It identifies the average treatment effect on the treated (ATT) under the (conditional) parallel trend assumption. Many studies use variation across groups that receive the treatment at different times similarly to the cycle superhighway programme. However, in the case of growing effects, using a classic regression with a treatment dummy and panel fixed effects does not recover a reasonable average of the treatment effects (Borusyak and Jaravel 2018). Concretely, in the case of increasing traffic, comparing different routes underestimates long-term effects as it evaluates lanes that just opened to lanes where traffic has been growing for a few years. Obtaining an average would confound the effect on lanes that have been opened for a year, three years, or more. Additionally, there is no meaningful "average treatment effect" as the different lanes' construction has happened at different times.

In more general terms, in the canonical DiD, the difference between the pre-treatment outcome is extrapolated to the post-treatment as a counterfactual. In a staggered DiD, the difference when the staggered groups have been treated also serves to identify the difference in level between groups. However, when the effects are not homogenous in time, then the $\hat{\beta}$ under-estimate long-term effects. Borusyak and Jaravel (2018) show that in the case where openings are distributed uniformly across time, the sample size weighted average treatment effect and the canonical regression estimand differs more and more as the effects become more dynamic (affecting the growth rate, see Figure 2.3.1 from Borusyak and Jaravel (2018)).

Figure 2.3.1: Difference in differences bias with dynamic effects




Another issue with the two-way DiD is that it overweights cohorts in the middle of the treatment and gives smaller weights to cohorts that opened first or last. A recent paper by Goodman-Bacon and Marcus (2020) shows that this estimator does not recover the ATT but a weighted estimator that depends on group size and variance in treatment. Again, it is mainly a problem if the effect changes over time (in the case of cycling lanes, increasing each year after opening). Other papers such as Borusyak and Jaravel (2018) describe this problem as "negative weighting" of the later cohorts.

Faced with this issue in the staggered difference in differences fixed-effect model, I use the programme's specificity to adopt an alternative identification strategy. I use an event study analysis to capture the impact of the opening of the lanes on cycling flows, car flows and accidents at various distances from the segregated lanes. The roll-out of the CSHs with segregated lanes is concentrated between 2015 and 2019, which allows me to estimate the effect up to 3 years after opening. I present the event study results with the constructed lanes only, as well as the lanes opened in 2020 as a control group. Assuming that the CSHs effect on traffic is stable across cohorts (they all receive the same impact at the opening and each year afterwards), then the event study estimates should be non-biased. In the summary statistics, I compute the socio-economic characteristics of the areas around the different cohorts of segregated lanes to look for differences that could impact the treatment. I do not find major differences in characteristics between cohorts.

Different endogeneity issues could arise in this setting. The first one would be that the timing of the opening is endogenous to the growth potential of the routes. Differences in the timing of opening were due to security concerns over the original design - it should not be related to the potential for cycle growth in the respective routes. The original plan was announced in 2008. It consisted of radial 12 roads linking to Central London.

The first lanes were built in 2010 and focused on improving the readability of the infrastructure. The lanes were visible using "blue paint" on the surface. They were not separated from the traffic and were perceived as unsafe. In response, TfL organised user consultations and small experimentations using the International Cycling Infrastructure Best Practice Study. The safety recommendations were integrated into the "Mayor's cycling vision" and led to higher safety standards. One major drawback of the higher standard of infrastructure was the substantial delays in implementing the CSH programme (Transport for London 2014). The second generation of lanes was physically separated from car traffic. It often involved reducing the number of car lanes to fit the 4 meters wide cycling lanes (compared to the non-separated 2m wide original design). There is no indication that the cycling potential was a factor in the timing of the roll-out.

The second endogeneity issue is that the routes could compete with each other for cyclists. The CSHs have been created to be radial roads spanning the London network. They are connecting different parts of London to the centres. The different routes are thus not substitutable. However, there might be a possibility that they are complementary - the more connected the network of segregated lanes, the more valuable they are for Londoners that can now travel safely for greater parts of their journey.

In all the regressions, I cluster the standard errors using two-way clustering at the CSHs route and year level. The general approach for an event study is to cluster at the unit or treatment level. If the error correlations are due to common shocks across observations, then the year-fixed effects will absorb all within-year clustering, and inference needs only to control for clustering on the unit. However, if these shocks have a large route-level component,

contemporaneous error correlations across routes will remain. I thus choose two-way clustering at the route and year level.

## 2.4 Data

I construct a dataset of cycling traffic flow, car traffic flow, and accidents for a representative set of counting sites along the Cycle Superhighway routes and their surrounding areas. Most of these points are located in Central London (inside the Congestion zone - where car traffic is tolled). For the displacement analysis, I consider locations close to these lanes (up to 600 meters).

*CSH data* - I use the Cycle Superhighway dataset from the Transport for London cycling monitoring programme. The counting sites are shown with the lines opening year in Figure 2.4.1. The CSH dataset has 320 count sites over 11 planned routes[5] [6]. For each site, I have their exact location and yearly count. These counts are based on daytime ridership and conducted annually; they are adjusted for seasonal variations and represent annual averages. It starts in 2014 and ends in 2019, but not all sites are surveyed every year: I have a balanced panel, pre and post-treatment counts for C1, CS2, CS3, CS5 and C6, which corresponds to 84 counting sites in the treatment.

*Cycle Hire* - To complement the survey analysis, I gather all cycle hire journeys[7] available from 2012 to today. I filter journeys corresponding to the same period (2014-2019) and time of the day (early morning to evening) as the survey data. I clean the data from all lost or incomplete journeys.

*London cycling*- I use the Central/Inner/Outer London Cycle Monitoring programme dataset for the cycle displacement analysis. It starts in 2014 and ends in 2019. Each counting site is monitored quarterly and observed in all directions. They are shown in Figure 2.4.2. For each site, I calculate the distance to the CSHs and group them by distance bands. This dataset has been sampled to be representative of London's cycling roads and traffic.

[5] I assign the route reference to counting sites based on the planned network map
[6] Each count site is observed in all directions
[7] They are also called Barclays bike or Boris Bike from the name of the first sponsor/ mayor of London

*Road traffic* - I use the Department for Transport Road Traffic Counts dataset for the car displacement analysis. It starts in 2002. It is observed yearly but with gaps. The road traffic data is available for every major road and some minor roads. The counters are represented in Figure 2.4.4. Similarly to the cycle analysis, I calculate the distance from each monitoring site to the CSH lines. As the data is not available every year, I use the provided imputed values for the missing years[8].

*Traffic accident* - Finally, I use the Road Safety Data (STATS19) available from 2004 to today from the Department for Transport to collect all incidents involving cars and bikes near the CSHs before and after opening. The data is precisely geocoded, which allows me to capture accidents on cycling lanes. The data only reports accidents with the police involved - it is thus likely to be missing non-serious accidents. There is no reason to think that the rate of reporting has changed over time. The data contains information on the severity of the accident. However, it is difficult to analyse by severity as the severity reporting was changed in 2016.

Table 2.4.1 shows the main census characteristics for a 150m buffer around the monitoring sites in my treatment groups. Columns 1, 2 and 3 correspond to routes opened in 2015, 2016 and 2018 respectively. As the number of monitoring sites observed for six years is low for the 2015 routes, I reproduce the results dropping that cohort and find similar results. The opening date does not correlate with demographic characteristics or total cycle traffic flow (in both directions). However, the earlier routes are a bit more central, leading to a slightly shorter travel time to work and a lower population. In Table 2.4.2, I present the same variables for the treatment groups (2015-2018) and the control group (never treated). The two groups' areas are similar in demographics and distance to work, but the treatment groups have a higher population overall and a slightly higher proportion of people biking to work.

[8]I do not allow imputation if the gap between two actual counts contains the year of construction

Table 2.4.1: Summary statistics 2011 census (treatment group)

| | 2015 | 2016 | 2018 |
|---|---|---|---|
| Household size | 1.9 | 2.22 | 1.95 |
| | (0.2) | (0.58) | (0.28) |
| Population | 268.2 | 360.47 | 336.94 |
| | (34.62) | (75.42) | (89.59) |
| Age | 39.71 | 33.02 | 35.57 |
| | (6.03) | (3.5) | (3.79) |
| Median age | 37.98 | 29.49 | 32.11 |
| | (6.65) | (4.98) | (4.92) |
| Share highly educated | 56.61 | 48.41 | 52.31 |
| | (2.31) | (18.34) | (14.17) |
| Bike to work (per 1000) | 40.08 | 35.34 | 29.94 |
| | (30.47) | (28.16) | (10.5) |
| Distance to work | 8.71 | 8.75 | 7.9 |
| | (2.21) | (2.52) | (1.68) |
| Total cycles | 2047.36 | 1425.74 | 1603.28 |
| | (830.95) | (1136.6) | (1019.97) |
| Counting sites # | N= 2 | N= 20 | N= 20 |

Table 2.4.2: Balance table 2011 census (treatment versus control)

| | Treated | Control | Treated=Control |
|---|---|---|---|
| Household size | 2.08 | 2.21 | p=0.24 |
| | (0.46) | (0.21) | |
| Population | 344.87 | 302.32 | p=0.02 |
| | (82.59) | (27.21) | |
| Age | 34.56 | 32.88 | p=0.24 |
| | (4.02) | (2.87) | |
| Median age | 31.14 | 30.31 | p=0.48 |
| | (5.28) | (1.93) | |
| Share highly educated | 50.66 | 41.89 | p=0.18 |
| | (15.95) | (13.09) | |
| Bike to work (per 1000) | 32.99 | 25.01 | p=0.08 |
| | (21.24) | (7.05) | |
| Distance to work | 8.34 | 8.56 | p=0.79 |
| | (2.13) | (1.71) | |
| Total cycles | 1539.89 | 1571.89 | p=0.91 |
| | (1057.59) | (540.27) | |
| Counting sites # | N= 42 | N= 6 | |

Figure 2.4.1: Map of the CSH counting sites

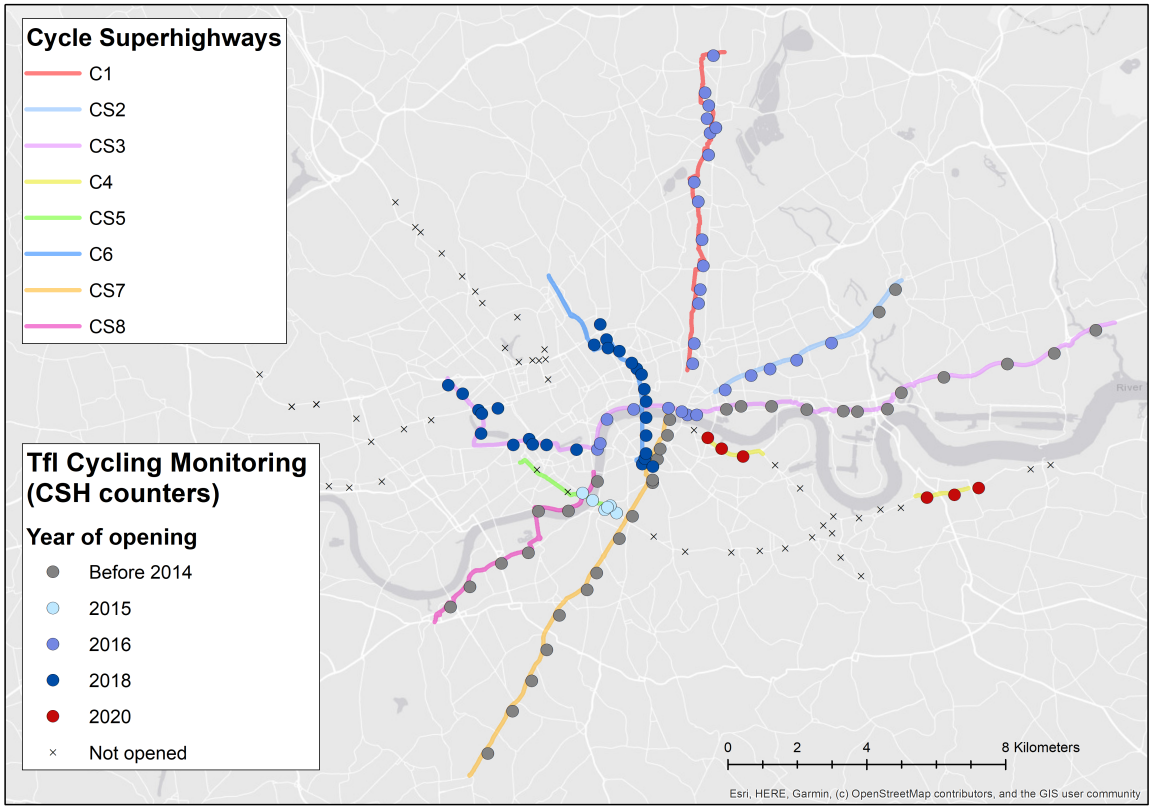



Figure 2.4.2: Map of the cycling monitoring programme counting sites

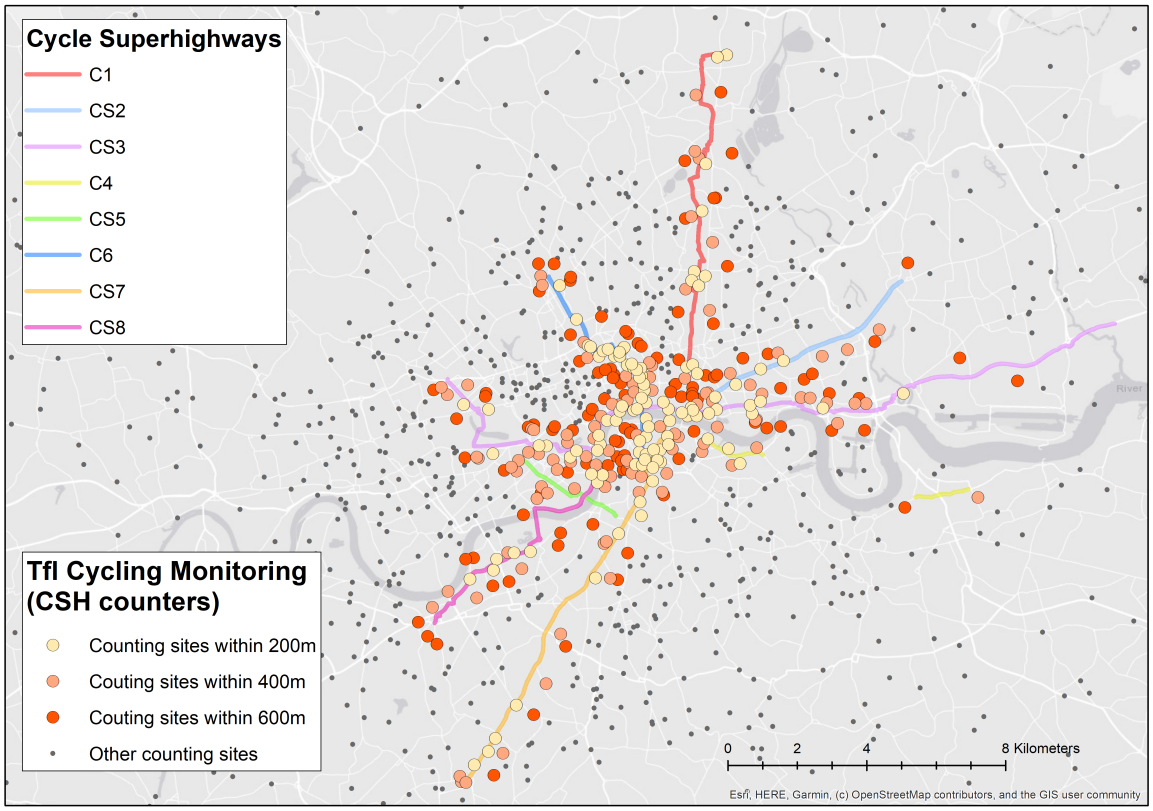


Figure 2.4.3: Map of the cycle hire stations

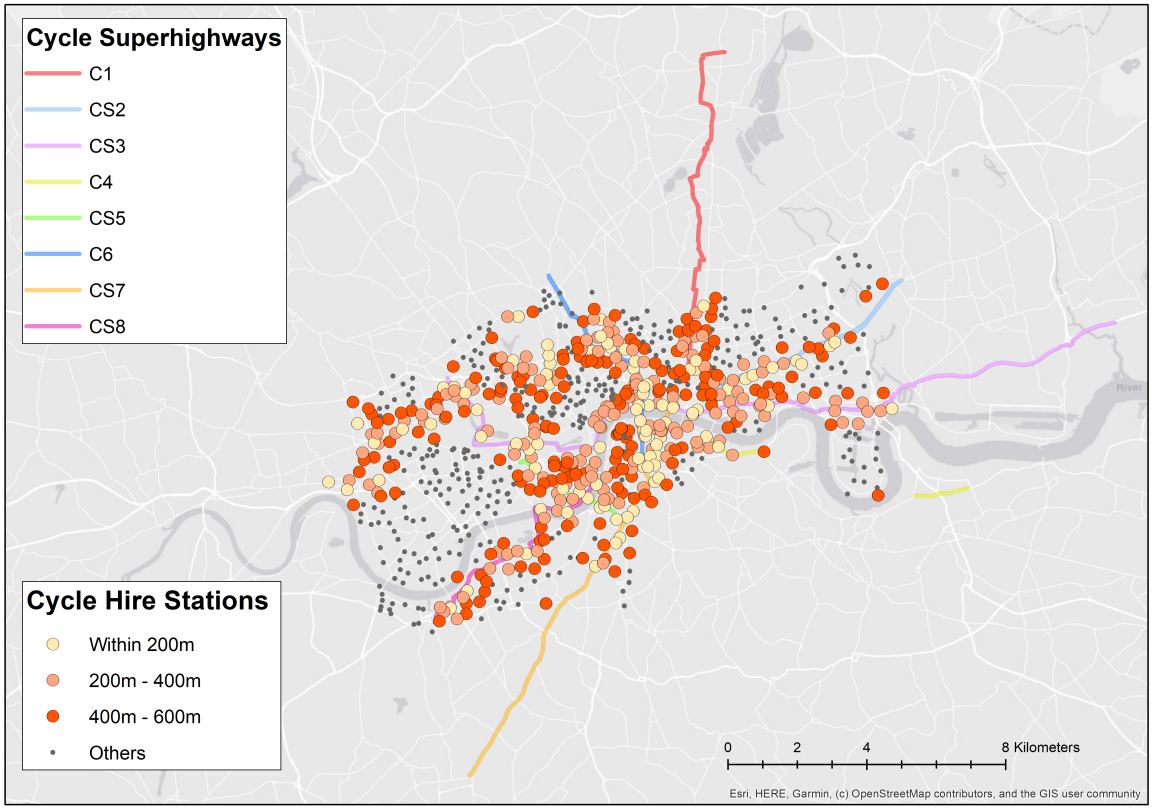



Figure 2.4.4: Map of road traffic counting sites

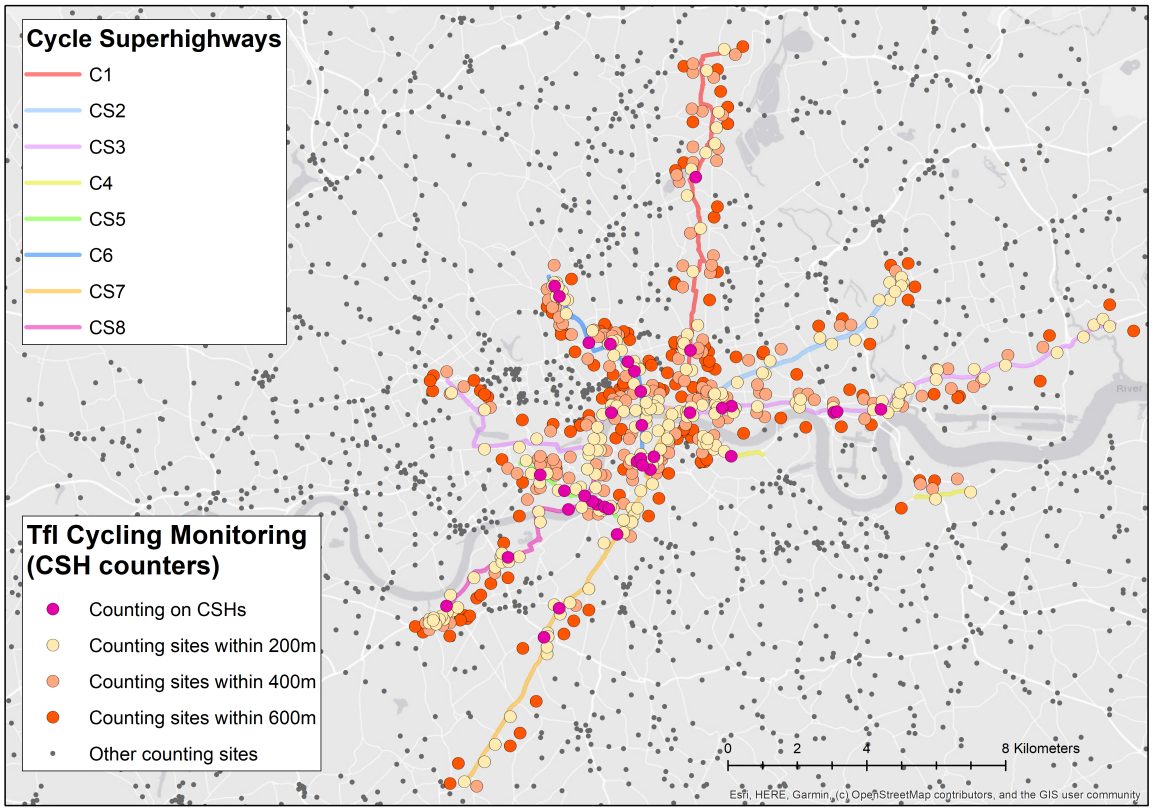
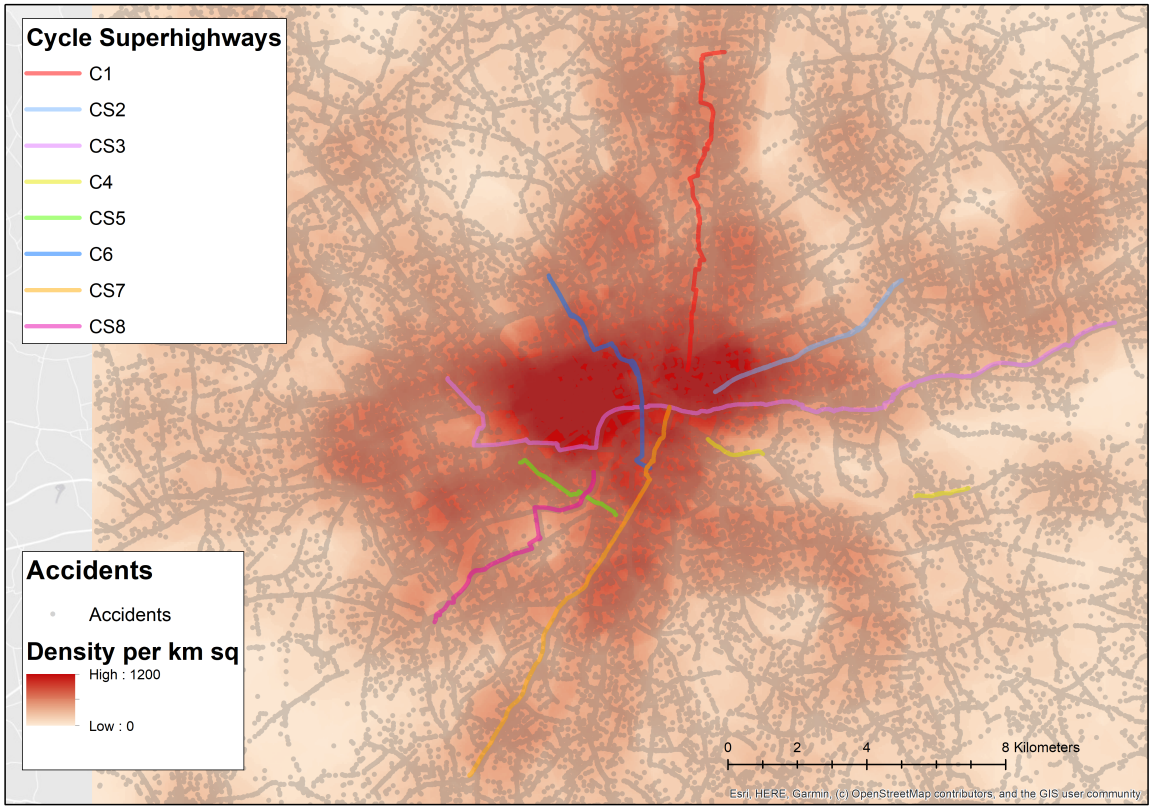


Figure 2.4.5: Map of accidents density 2009-2019



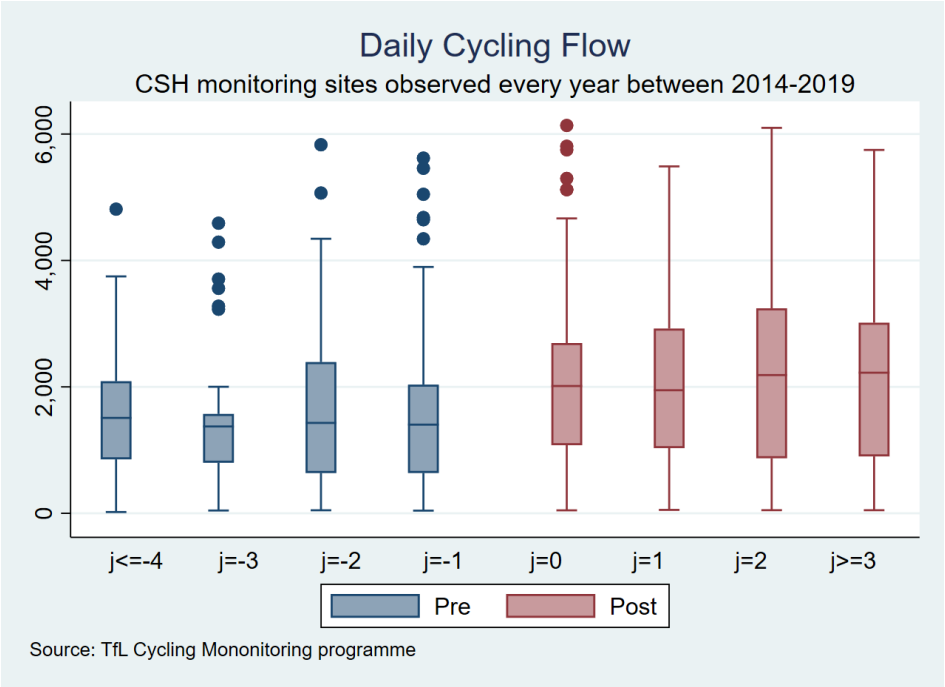

## 2.5 Empirical analysis and results

To introduce the analysis, Figure 2.5.1 plots the raw count of the average daily cycling flow for the treated sample for each year pre and post-treatment. The average pre-treatment is about 1,600 counted daily in the monitoring sites, against 2,100 post-treatment. The lower and upper whiskers show that some counting stations average low numbers daily (the minimum is 20 pre-treatment and 46 post-treatment while the busiest monitoring stations register up to 5,830 and 6,136 cyclists counted daily pre-and post-treatment, respectively). These results do not consider the overall growth trends and site heterogeneity but still show a significant jump at opening.

### 2.5.1 Dynamic effects estimation

The results from Figure 2.5.1 indicate that it would be appropriate to conduct an event study on the CSH openings. I use a fully dynamic specification to analyse the treatment heterogeneity

Figure 2.5.1: Daily cycling flow




across time. Unlike the two-way fixed effect DiD, the event study coefficient should not be affected by a negative bias as long as the increase is similar across cohorts (cohorts receive the same increase at the opening and every year afterwards). I only include sites for which I have six years of observation in the sample. The treatment group is CSH sites after opening in 2015, 2016 and 2018. The control group is sites that have not opened yet (later treated) and CSH sites that opened in 2020. The dependent variable is the log flow of cyclists.

$$\ln(TotalCycles_{it}) = \sum_{j=-4}^{J} \theta^j Treat_{it}^j + \gamma_i + \delta_t + \eta_{it}$$

with $ln(TotalCycles)_{it}$ the average daily flow recorded in counter $i$ and year $t$. As I use a log-linear model and the coefficients for years of opening are quite large, I exponentiate them in the text. $Treat_{it}^j = \mathbb{1}\{j = t - Opening_i\}$ is a categorical variable for years since opening $Opening_i$ $j = \{-3, -2, ..., 4\}$: I use $j = -1$, the year before opening, as a base level, $\theta_j$ for $j \geq 0$ captures dynamic effects of $j$ years relative the cycle superhighway opening, finally $\gamma_i$ and $\delta_t$ site and year fixed effects.

I show the result in Table 2.5.1. In column 1, I only include sites for which I have pre and post-treatment years and six years of observation. In column 2, I add the CSH that

opened in 2020 (I do not observe the flows after opening as my panel stops in 2019). The coefficients are slightly larger than in column 1 but still within the confidence interval of each other. I illustrate the results of column 1 in the upper left graphs of Figure 2.5.2. There is a short decrease before opening, probably due to construction, a large jump of 24% at the opening, and then a further average increase of 19% per additional year. From the author's calculation, the average construction time is 14 months, which can explain the slight decrease up to two years before opening[9]. The estimates with the control group are a bit larger but also less precisely estimated. I also estimate an average treatment effect on the treated using a difference in differences estimator which I show in the Appendix in Table 2.A.1 and the corresponding Goodman-Bacon decomposition in Figure 2.A.4. Both confirm that there is a large increase after the opening of the CSHs. The effects are increasing over time.

### 2.5.2 Traffic displacement

In the next part of the analysis, I reproduce the event study on traffic flow for cycles and cars close to the newly constructed CSH. The estimating equation is the same as above but uses cycle traffic around the CSH as an outcome. To this aim, I use the Cycle monitoring dataset from Transport for London for Central, Inner, and Outer London presented in Figure 2.4.2. I keep all counting sites opened between 2015 and 2020 for the analysis and use the year before opening as the baseline.

The results from the event study on cycle counters 20-200m, 200-400m, and 400-600m away are presented in Table 2.5.2 and in the last 3 plots of Figure 2.5.2. I keep all counting sites that I observe for all quarters. While the coefficients after opening are larger closer to CSHs - meaning that the cycling traffic could be increased close to the CSHs, there are no significant results. The effects could be linked to cyclists getting on and off the segregated lanes. On average, cycle trips are short (20min) and fast, so there is not much gain for the average cyclist to take a large detour to get on a cycle superhighway.

[9]CS3 from Tower Gateway to Parliament Square took 13 months to be built and opened in March 2016. CS3 from Parliament Square to Lancaster gate started in April 2016 and ended in September 2018 (18 months). CS5 Kennington Lane to Victoria took eight months. CS1 took eight months between was built between July 2015 and April 2016. CS6 started in March 2015 and finished in September 2018. CS2 extension started in February 2015 and ended in December 2016 (21 months)

Table 2.5.1: Cycling flow after CSH opening

| | Treated | Treated + Control |
|---|---|---|
| j<=-4 | -0.309 | -0.299 |
| | (0.174) | (0.212) |
| j=-3 | -0.118 | -0.126 |
| | (0.0767) | (0.149) |
| j=-2 | -0.166*** | -0.135** |
| | (0.0322) | (0.0447) |
| j=0 | 0.215*** | 0.260*** |
| | (0.0397) | (0.0315) |
| j=1 | 0.345*** | 0.400*** |
| | (0.0208) | (0.0283) |
| j=2 | 0.494*** | 0.562*** |
| | (0.0416) | (0.104) |
| j>=3 | 0.595*** | 0.696*** |
| | (0.0185) | (0.105) |
| N | 504 | 528 |
| Rsquared | 0.949 | 0.949 |
| Year FE | Yes | Yes |
| Site FE | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

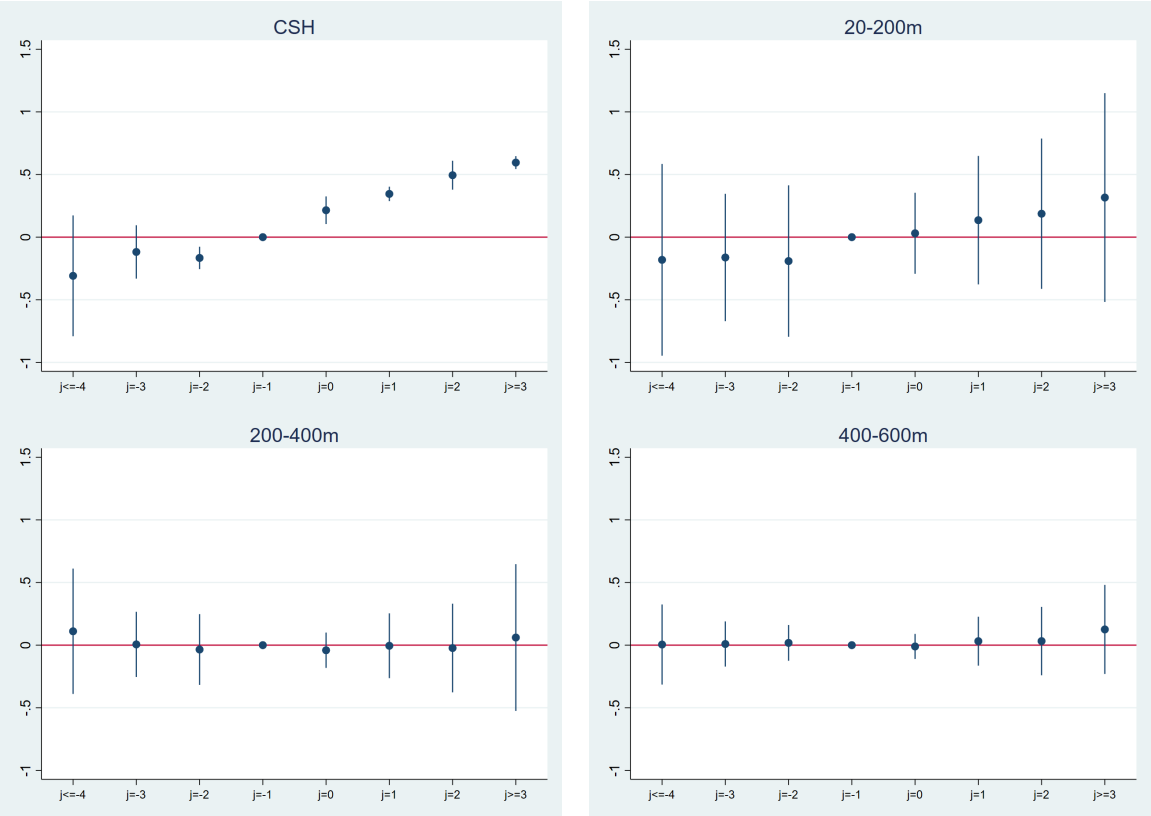
Table 2.5.2: Cycle displacement

| | 20-200m | 200-400m | 400-600m |
|---|---|---|---|
| j<=-4 | -0.181 | 0.111 | 0.00487 |
| | (0.178) | (0.157) | (0.115) |
| j=-3 | -0.163 | 0.00582 | 0.00909 |
| | (0.118) | (0.0818) | (0.0649) |
| j=-2 | -0.191 | -0.0350 | 0.0180 |
| | (0.141) | (0.0888) | (0.0514) |
| j=0 | 0.0308 | -0.0408 | -0.0105 |
| | (0.0752) | (0.0443) | (0.0360) |
| j=1 | 0.135 | -0.00498 | 0.0316 |
| | (0.119) | (0.0813) | (0.0703) |
| j=2 | 0.187 | -0.0230 | 0.0323 |
| | (0.139) | (0.111) | (0.0984) |
| j>=3 | 0.316 | 0.0608 | 0.125 |
| | (0.194) | (0.184) | (0.128) |
| N | 1426 | 2415 | 3151 |
| Rsquared | 0.898 | 0.909 | 0.932 |
| Quarter FE | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes |

SD clustered at year and cycle superhighway route level

$^{*}$ $p < .1$, $^{**}$ $p < .05$, $^{***}$ $p < .01$

Figure 2.5.2: Cycling flows near CSHs using TfL cycling surveys




A drawback of the Transport for London cycling surveys is that there are only conducted annually. That is why I also do a robustness check on the impact of the segregated lanes on cycling using the cycle hires data provided by Transport for London. The dataset has all journeys done by hire bikes in London from 2012 to March 2020 (more recent data is available, but I wanted to exclude any changes due to lockdowns). I restrict the analysis to the stations near lanes opened after 2014. The dependent variable is the logged number of journeys starting or ending near segregated lanes. I subset the sample to stations on the segregated lanes (0-20m) and then 20-200m, 200-400m and 400-600m away. I present the results in the appendix in Table 2.A.4 and 2.A.5 , and graphically in Figure 2.5.3 and 2.5.4. I find the same increase in hire starting or ending near segregated lanes, but the effect disappears for stations more than 200m away.

To rule out that I am capturing the impact of contemporaneous policies, I look at possible links with other transport and cycling policies. The two other major cycling policies happening during the same period are the Biking Boroughs project and the roll-out of the London Cycle

Figure 2.5.3: Cycle hire journeys starting near CSHs

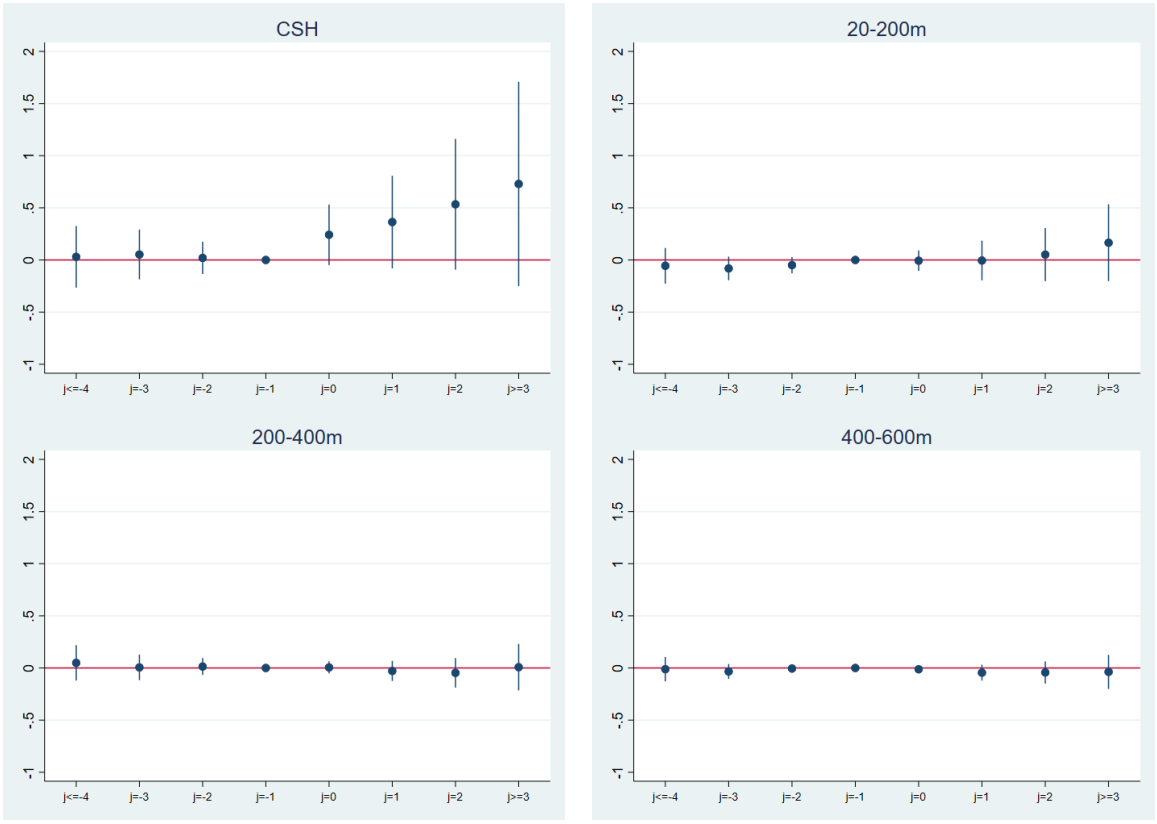
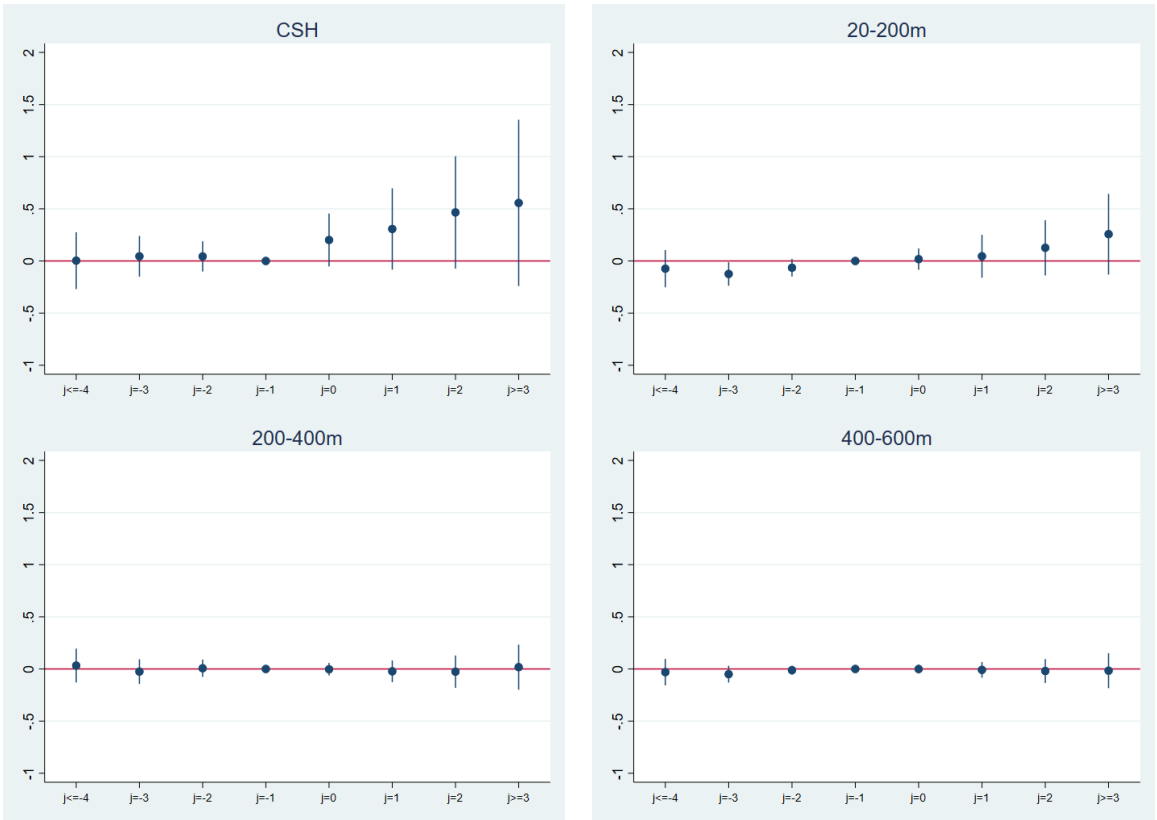



Figure 2.5.4: Cycle hire journeys ending near CSHs



Hire Scheme. The Biking Borough project was aimed at Outer London Boroughs - it is thus a different geographical zone to the CSHs which are concentrated in Central London. I can exclude that they have a direct impact on the CSHs.

The London Cycle Hire Scheme opened in 2010 with a second major extension in 2012. Most of their implementation is thus prior to the development of the segregated lanes. 97% of my CSH survey points have a cycle hire station within 200m. I plot an histogram of the date of opening of the cycle hire stations within 200m of the CSHs compared to the opening of the cycling lanes in Appendix 2.A.5. I find that 75% of the cycle hire stations opened before the CSHs' openings. There is a clear stop of two years during the CSHs' constructions and then after opening, the implementation continues on the same decreasing trend. To further rule out that cycle hire stations' openings contribute significantly to the increase of cycling flows after the segregated lanes opening, I analyse the impact of getting a new cycle hire station within 200m of the segregated lanes. I find no significant impact of getting a new station in appendix Table 2.A.6. I cannot do an event study for the opening of the stations as I do not observe enough openings.

The CSHs and the cycling hire are two very effective policies for increasing cycling flows. Clearly, neither the placement of the cycle hire stations nor the placement of the segregated lanes was random. They have both been selected to bolster cycling usage on roads with high potential. However, I argue that the increase that I observe in the CSHs event analysis is solely due to the segregation of the lanes as I do not observe any pre-trends in the CSHs and the cycling hire analysis. Getting additional stations does not increase massively the traffic on the segregated lanes - probably because most of the stations were already built by the time the lanes got constructed. The pre-opening levels, however, probably reflect the already high flows of these lanes and the impact of the cycle hire scheme.

I then reproduce the same event study for cars' and buses' displacement. The outcome is the logged number of total cars (or buses) counted at each survey point. To this aim, I use the yearly counts provided by the Department for Transport. The dataset is available for all of England, but I concentrate on counters on a CSH route (road segment where the lanes were

built by reducing car lane capacity, 0-20m) or close to it (20-200m, 200-400m and 400-600m distance buffers). The car traffic dataset provided by TfL relies on imputed data (all counters are not observed each year and the data is interpolated from past years). I present the results using an unbalanced panel in Table 2.5.3 and Figure 2.5.5. I also show the results for the interpolated balanced panel in the appendix Table 2.A.7 for cars and in appendix Table 2.A.8 for buses.

Both tables are fairly similar. The coefficients are slightly higher and positive on the roads with the CSH but not statistically significant.

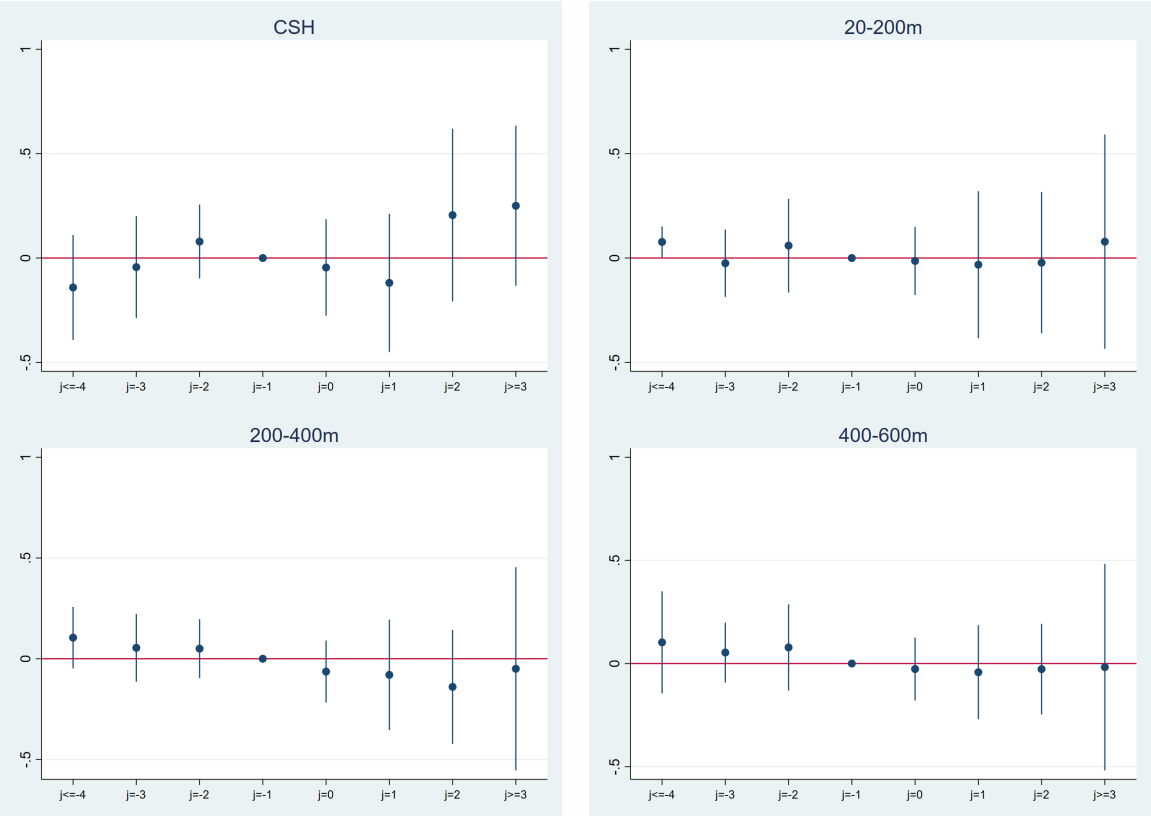
Table 2.5.3: Car displacement near CSH

| | CSH | 20-200m | 200-400m | 400-600m |
|---|---|---|---|---|
| j<=-4 | -0.141 | 0.0770** | 0.105 | 0.103 |
| | (0.0905) | (0.0288) | (0.0593) | (0.0964) |
| j=-3 | -0.0437 | -0.0251 | 0.0543 | 0.0535 |
| | (0.0879) | (0.0629) | (0.0653) | (0.0566) |
| j=-2 | 0.0787 | 0.0593 | 0.0501 | 0.0781 |
| | (0.0638) | (0.0873) | (0.0569) | (0.0813) |
| j=0 | -0.0456 | -0.0139 | -0.0637 | -0.0266 |
| | (0.0832) | (0.0634) | (0.0600) | (0.0592) |
| j=1 | -0.119 | -0.0319 | -0.0800 | -0.0422 |
| | (0.119) | (0.137) | (0.107) | (0.0887) |
| j=2 | 0.205 | -0.0224 | -0.140 | -0.0271 |
| | (0.149) | (0.132) | (0.110) | (0.0854) |
| j>=3 | 0.250 | 0.0783 | -0.0498 | -0.0175 |
| | (0.138) | (0.200) | (0.196) | (0.195) |
| N | 212 | 510 | 782 | 988 |
| Rsquared | 0.974 | 0.967 | 0.968 | 0.983 |
| Year FE | Yes | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Figure 2.5.5: Cars flows near CSHs




### 2.5.3 Accidents reduction

In the last part of the analysis, I look at accident reduction on CSHs. First, I use the STATS 19 datasets that record the location of traffic accidents involving the police in England. Next, I add traffic flow for the average DfT counting sites on the CSHs. The set of monitoring points with traffic data is small, but the results are consistent across specifications.

I include all accidents located on CSH lanes constructed after 2014. In Table 2.5.4 and Figure 2.5.6, I look at the difference between painted lanes and lanes fully segregated by a kerb (the car traffic is physically separated from the cycling lanes). The reduction in accidents seems to be driven by the latter, even though a small sample size might also be at play here.

In Table 2.5.5, I present the results for total accidents involving cyclists, total accidents involving cyclists divided by cycling flow, total accidents involving cyclists divided by cars' flow and total accidents involving cars divided by cars' flow.

In column 1, there is a significant decrease in total accidents after a CSH opening. The results even hold when looking at the number of cycling accidents per cyclist in column 2. In line with the literature, these results indicate that separating cyclists from motorised traffic reduces the number of accidents. Both the number of accidents per cyclist and the absolute number of accidents decrease, indicating that there is safety in numbers - cars are more likely to expect cyclists if they see cycling infrastructures.

In column 3, I look at cycle accidents by car flow. The coefficients become negative after the lanes' opening - but they are not significant at the standard significance level. In columns 4 and 5, I look at car accidents after the opening of the segregated lane. There is no significant pattern emerging. The lanes do not seem to have made traffic safer for cars.

Figure 2.5.6: Bike accident after CSH opening

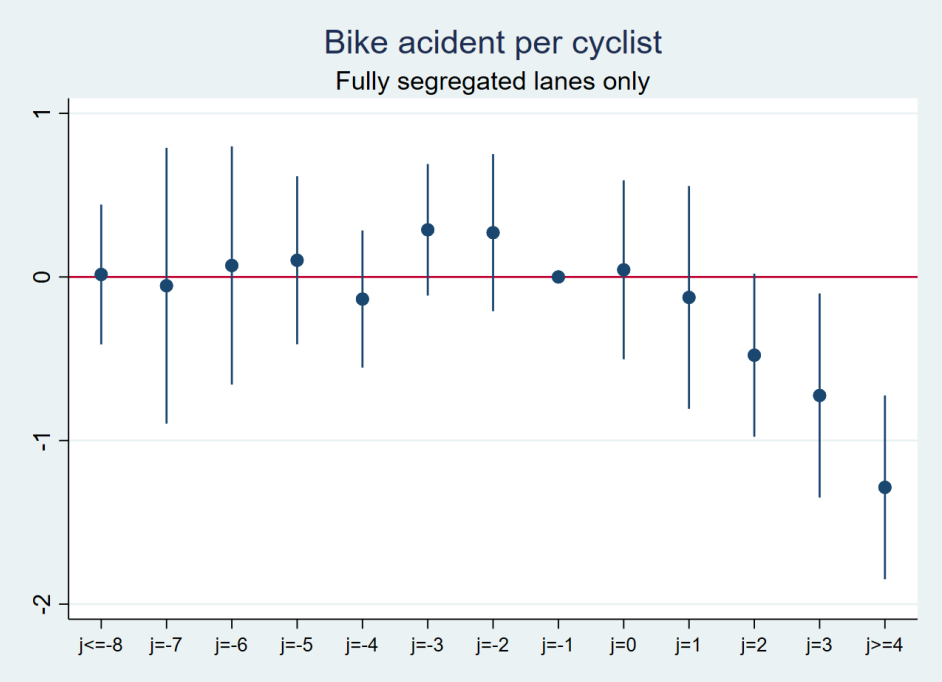


Table 2.5.4: Bike accidents over bike flow by lane segregation

| | Painted<br>Bike acc./bike flow | Fully segregated<br>Bike acc./bike flow |
|---|---|---|
| j<=-8 | 2.625 | 0.0152 |
| | (1.718) | (0.154) |
| j=-7 | 2.046 | -0.0536 |
| | (1.554) | (0.304) |
| j=-6 | 1.539 | 0.0701 |
| | (1.068) | (0.262) |
| j=-5 | 0.406 | 0.102 |
| | (1.094) | (0.185) |
| j=-4 | 1.137 | -0.135 |
| | (0.647) | (0.151) |
| j=-3 | -0.205 | 0.288 |
| | (0.720) | (0.145) |
| j=-2 | -0.436 | 0.271 |
| | (0.742) | (0.173) |
| j=0 | -1.673 | 0.0439 |
| | (0.905) | (0.197) |
| j=1 | -0.891 | -0.125 |
| | (0.826) | (0.245) |
| j=2 | -2.460 | -0.478* |
| | (1.161) | (0.180) |
| j=3 | -1.455 | -0.725** |
| | (0.696) | (0.225) |
| j>=4 | | -1.286*** |
| | | (0.202) |
| N | 55 | 154 |
| Rsquared | 0.692 | 0.811 |
| Year FE | Yes | Yes |
| Site FE | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2.5.5: Traffic accidents

| | Bike acc. | Bike acc. vs bike flow | Bike acc. vs car flow | Car acc. | Car acc. vs car flow |
|---|---|---|---|---|---|
| j<=-8 | 1.350 | 0.808 | 0.505 | -0.613 | -0.916 |
| | (0.975) | (1.066) | (0.852) | (0.706) | (0.580) |
| j=-7 | 1.071 | 0.651 | 0.415 | -0.653 | -0.889* |
| | (0.710) | (0.838) | (0.699) | (0.480) | (0.390) |
| j=-6 | 0.834 | 0.460 | 0.260 | -0.419 | -0.619 |
| | (0.542) | (0.483) | (0.307) | (0.422) | (0.336) |
| j=-5 | 0.611 | 0.286 | 0.125 | -0.678* | -0.838** |
| | (0.418) | (0.540) | (0.452) | (0.253) | (0.190) |
| j=-4 | 0.504 | 0.209 | 0.0653 | -0.364 | -0.507 |
| | (0.356) | (0.348) | (0.268) | (0.442) | (0.405) |
| j=-3 | 0.324 | 0.0360 | -0.0611 | -1.033* | -1.130** |
| | (0.177) | (0.257) | (0.244) | (0.401) | (0.366) |
| j=-2 | 0.264 | 0.128 | 0.124 | -0.0817 | -0.0863 |
| | (0.160) | (0.188) | (0.202) | (0.161) | (0.166) |
| j=0 | -0.268** | -0.280** | -0.146 | -0.263 | -0.129 |
| | (0.0871) | (0.0965) | (0.0943) | (0.343) | (0.315) |
| j=1 | -0.545 | -0.380 | -0.148 | -0.108 | 0.123 |
| | (0.356) | (0.444) | (0.508) | (0.365) | (0.326) |
| j=2 | -0.995** | -0.789** | -0.497 | -0.310 | -0.0175 |
| | (0.224) | (0.242) | (0.343) | (0.539) | (0.398) |
| j=3 | -1.283** | -0.875* | -0.551 | 0.153 | 0.477 |
| | (0.348) | (0.407) | (0.590) | (0.596) | (0.471) |
| j>=4 | -1.978*** | -1.401** | -1.227* | 0.250 | 0.423 |
| | (0.354) | (0.431) | (0.525) | (0.598) | (0.467) |
| N | 209 | 209 | 209 | 209 | 209 |
| Rsquared | 0.751 | 0.715 | 0.723 | 0.609 | 0.617 |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

## 2.6 Conclusion

The cycle superhighway programme is associated with a large increase in cycle traffic. The treatment effect at opening represents an increase of about 25% in ridership. The effect increases over time by about 20% a year. Most of this increase is due to new cyclists and increased cycling frequency as there is no evidence of cyclist displacement or car traffic displacement. One of the factors investigated in this analysis to explain the rise in ridership is an increase in safety due to a larger number of cyclists and safer lanes. These findings are essential for policymakers as they show that infrastructures like cycling lanes should not be evaluated by the immediate impact but also by the continuous growth after opening. The agglomeration effects of the lanes are an essential factor to consider - the more lanes, the more cyclists, the safer they are, and the more likely people will take up cycling.

These results are essential to justify the construction of segregated lanes on major roads to encourage cycling. Moreover, in cities like London, where one of the main obstacles to cycling is safety perception, cycling lanes are essential to convince people to take up cycling. TfL surveys show indeed that new cyclists - for example, new e-bikes users - are particularly sensitive to these infrastructures as they provide safety and clear directions to connect to central parts of the city. The lanes provide a good infrastructure start to sustain cycling growth.

An interesting alley for research would be to investigate how the connectivity and the spread of these lanes participate in increasing cycling usage. The ability to reach most of Copenhagen or Berlin via safe cycling paths is essential to their success. The current growth of the network offers an opportunity to study this phenomenon as it develops. The new "cycleways" programme launched in late 2019 aims to unify London's cycling projects (CSHs network, quiet ways and mini-hollands) to provide the best cycling routes between key destinations as part of a connected and unified network. It is an exciting venue for future research on cycling networks.

# Appendix

## 2.A.1 Illustration of the first and second phases of the cycle superhighways

Figure 2.A.1: Original cycle superhighways network map in 2009

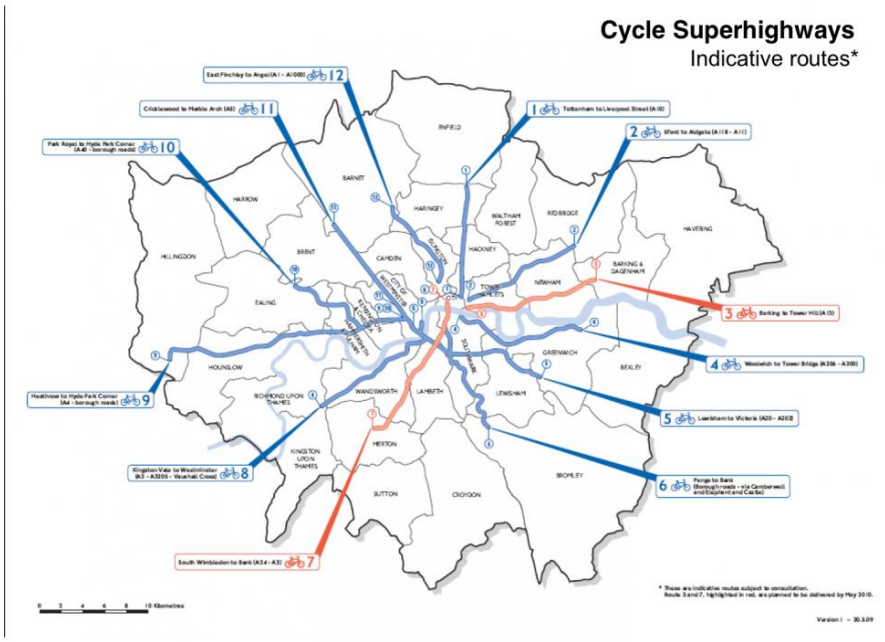



Figure 2.A.2: Cycle Superhighway 8 - Opened in 2011 - Painted lane

Figure 2.A.3: Cycle Superhighway 5 - Opened in 2015 - Segregated lane

### 2.A.2 Difference in differences

In this appendix, I present the difference in differences approach and the Goodman-Bacon decomposition on cycling flows.

I report the results for the OLS and the difference in differences using two ways fixed effect in Table 2.A.1. I only include sites for which I have six years of observation in my sample. The treatment group is CSH sites that opened in 2015, 2016 and 2018. The control group is CSH sites that were planned but not opened yet and sites that opened in 2020.

The dependent variable is the flow of cyclists logged, I interpret the coefficient on the opening of the segregated lane $CSH_{i,t}$ as the variation in the percentage of the conditional mean of the regressand. As the coefficients are quite large, I exponentiate them in the text. In columns 1, and 2, I only use only treated sites (the ones opened between 2015 and 2018). In columns 3 and 4, I add the routes or part of routes that have been planned but not built and

the ones that have been built later in 2020 [10]. In the OLS estimation, I control for the local borough.

In column 1, the naive OLS effect is quite large, a 48% increase, but could be suffering from bias from differences between sites. The impact of getting a CSH is reduced to 17% once site fixed effects are included; however, this specification is likely to suffer from significant bias in a staggered setting; it is a weighted average of the different lengths of exposure with a downward bias as it compares late to early treated.

In column 3, I add a control group using the route opened in 2020 and never constructed. Introducing a control group allows the bias introduced by the late treated to the early treated to be slightly reduced. While the OLS results between columns 1 and 3 are similar, the coefficient for the two ways FE in column 4 is closer now to 30% compared to 17% without the control group. It is consistent with the two-way FE DiD estimator being biased in case of increasing heterogeneous effect in time.

Table 2.A.1: OLS and FE estimations

| | Treated | | Treated + Control | |
|---|---|---|---|---|
| | OLS | FE | OLS | FE |
| Post | 0.394*** | 0.159 | 0.357** | 0.264* |
| | (0.0833) | (0.0892) | (0.102) | (0.105) |
| N | 504 | 504 | 576 | 576 |
| Rsquared | 0.650 | 0.948 | 0.637 | 0.945 |
| Year FE | No | Yes | No | Yes |
| Site FE | No | Yes | No | Yes |
| Controls | Yes | No | Yes | No |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

I then present the results of the decomposition of the difference in differences fixed effects in a staggered setting.

The first comparison group is easily understandable: for each cohort, it compares the

[10] CS9, CS10 and CS11 were planned but were not constructed as of 2019 and CS9 opened in 2020, part of CS4 and CS5 did not get constructed; I never include routes that were constructed before 2014

treated cohorts (e.g. opening in 2015, 2016, 2018) with the control group (routes that opened in 2020 or were never opened). As long as the control group is a good counterfactual for the treated groups, these differences capture the impact of the segregated lanes. The estimates are represented by grey triangles in the graph.

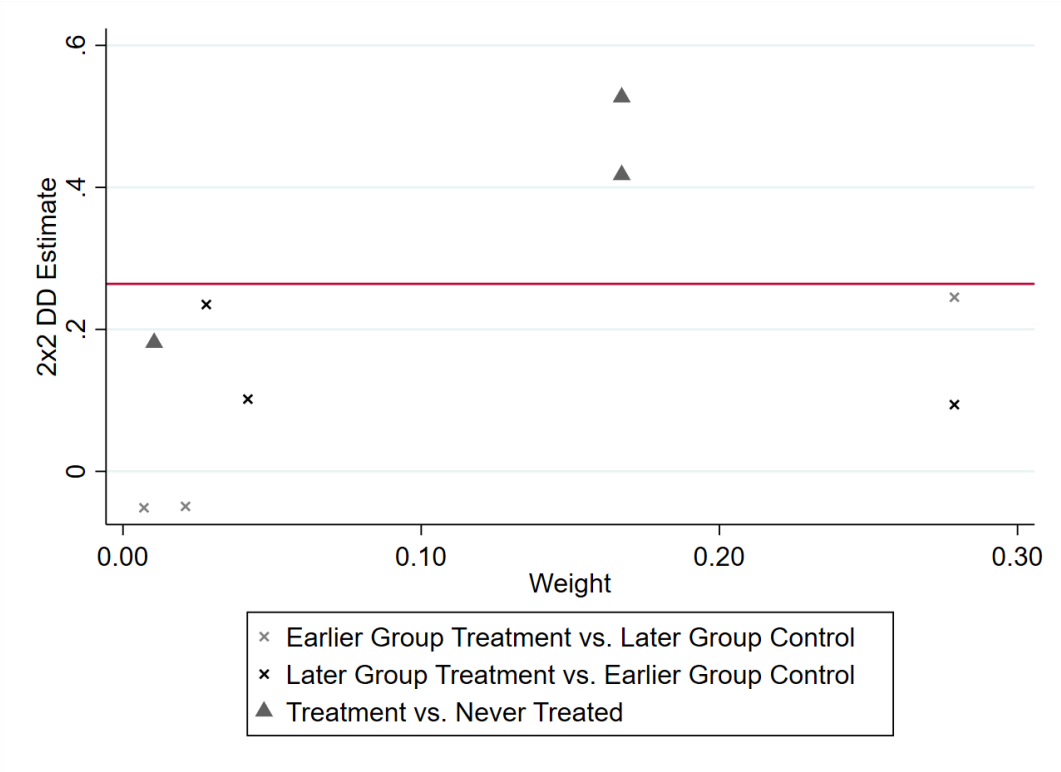
The next comparison - early treated versus late as control is also fairly straightforward as long as there is no anticipation of the treatment. For example, it compares the sites opened in 2015 with sites opened later but before they were opened. We thus get three comparisons: 2015 with 2016, 2015 with 2018 and 2016 with 2018. The estimates are represented by grey crosses on the graph.

The last set of comparisons is the problematic one. They compare early treatment versus late as control (after late gets treated). It assumes that the pre-treatment difference should be equal to the post-treatment. But if there is maturation in treatment (as in the case of the cycling lanes), the after gap is likely to be larger as the early treated have more time to grow, and it will overall under-estimate the treatment effect. In general, treatment effects change (monotonically) over time, the DiD estimate is biased away from the sign of the true effects. The coefficients are represented by the black crosses.

I present the result for cycling flows on CSHs in Figure 2.A.4. The red line corresponds to the DiD estimator of Table 2.A.1 column 5. The overall estimator for earlier group versus later group control is 0.22 (cohorts opened in 2015 vs 2016 and 2018, and 2016 against 2018). The black crosses correspond to the latter group as treatment versus the earlier group as control after opening. The Later treatment vs Earlier control overall estimator is only 0.11. It is the estimator likely to be biased in case of increasing treatment effect over time. Finally, the treatment versus never treated is represented by the triangles. The corresponding coefficient is 0.46 (cohorts opened in 2015, 2016 and 2018 vs 2020 and never opened).

The x-axis in Figure 2.A.4 shows the weight allocated to each comparison based on group size and time in treatment.

Figure 2.A.4: Goodman-Bacon Decomposition




These figures indicate that the CSHs have been successful in attracting users after opening. I do a robustness check by checking the degree of segregation around the counting sites using the first difference approach. One of the main criticisms of the CSH scheme is that parts of the lanes are not fully segregated by a kerb, but only painted in blue, sometimes with bollards to delineate their locations. Using the London's Cycling Infrastructure Database (CID) created in 2018, I look at the impact of the degree of segregation of the lanes. In Table 2.A.2, I reproduce the DiD two-ways fixed effects estimate of Table 2.A.1 column 4 using not yet treated or never treated as a control. I find in column 1 that painted only lanes still see a large increase in cycling traffic, but the effect is only significant for fully segregated lanes.

I also repeat the event study analysis using the fully segregated lanes only (I can not do it on the painted lanes only, there are not enough observations for each year after treatment). I find similar coefficients than on the full sample - slightly lower for the last two years. These results seem to indicate that the increase in traffic flow is not only driven by the full segregation but also by other factors brought by the programme such as better visibility of cyclists and network effects.

Table 2.A.2: Lane segregation

| | Painted | Fully segregated |
|---|---|---|
| Post | 0.256 | 0.278* |
| | (0.151) | (0.127) |
| N | 276 | 300 |
| Rsquared | 0.971 | 0.903 |
| Year FE | Yes | Yes |
| Site FE | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2.A.3: Cycling flow after CSH opening by segregation

| | All | Fully segregated only |
|---|---|---|
| j<=-4 | -0.309 | -0.333 |
| | (0.174) | (0.171) |
| j=-3 | -0.118 | -0.0512 |
| | (0.0767) | (0.0661) |
| j=-2 | -0.166*** | -0.137* |
| | (0.0322) | (0.0519) |
| j=0 | 0.215*** | 0.262*** |
| | (0.0397) | (0.0532) |
| j=1 | 0.345*** | 0.290*** |
| | (0.0208) | (0.0530) |
| j=2 | 0.494*** | 0.370*** |
| | (0.0416) | (0.0716) |
| j>=3 | 0.595*** | 0.441*** |
| | (0.0185) | (0.0189) |
| N | 504 | 264 |
| Rsquared | 0.949 | 0.905 |
| Year FE | Yes | Yes |
| Site FE | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

### 2.A.3 Cycle hire analysis

I also do a robustness check of the impact of the segregated lanes on cycling using the cycle hires data provided by Transport for London. I use the same event study approach. The dataset has all journeys done by hire bikes in London from 2012 to March 2020 (more recent data is available but I wanted to exclude any changes due to lockdowns). I restrict the analysis to the lanes opened after 2014. The dependent variable is the logged number of journeys starting or ending near segregated lanes. I subset my sample to stations on the segregated lanes and then 200m, and 400m away. Contrary to the counting sites analysis, I do not know if the cyclists have used the segregated lanes, only that the journeys have started or ended near a segregated lane.

I find a similar (but less significant) increase on the segregated lanes (within 20 meters) but no effect further away. The standard errors for the groups further away are quite small, which gives confidence that the absence of displacement is real and not due to a lack of data.

### 2.A.4 Car displacement

I show below the event study for car and bus displacement after the opening of the lanes. The dataset of car counts in London provided by Transport for London uses a large number of imputed values. I remove all values where the imputation happens at the opening of the cycle lanes, and I only keep the values where I have a count before and after treatment. I present in Table 2.A.7 the results. The outcome is the logged number of cars or buses observed at each counting site on a typical day. The counting sites are observed every quarter. The dependent variable is the number of years before and after opening. The base level is the year before opening. As for the unbalanced panel, there is no evidence that cars' or buses flows have changed after the opening of the segregated lanes.

Table 2.A.4: Cycle hire journeys starting near CSH

| | CSH | 20-200m | 200-400m | 200-600m |
|---|---|---|---|---|
| j<=-4 | 0.0303 | -0.0557 | 0.0486 | -0.0121 |
| | (0.139) | (0.0841) | (0.0835) | (0.0578) |
| j=-3 | 0.0525 | -0.0813 | 0.00538 | -0.0335 |
| | (0.112) | (0.0556) | (0.0604) | (0.0351) |
| j=-2 | 0.0197 | -0.0494 | 0.0138 | -0.00541 |
| | (0.0731) | (0.0386) | (0.0398) | (0.0197) |
| j=0 | 0.241* | -0.00690 | 0.00545 | -0.0127 |
| | (0.137) | (0.0485) | (0.0284) | (0.0192) |
| j=1 | 0.364 | -0.00550 | -0.0287 | -0.0451 |
| | (0.209) | (0.0937) | (0.0476) | (0.0373) |
| j=2 | 0.534* | 0.0515 | -0.0468 | -0.0434 |
| | (0.296) | (0.125) | (0.0694) | (0.0522) |
| j>=3 | 0.729 | 0.165 | 0.00747 | -0.0374 |
| | (0.462) | (0.181) | (0.109) | (0.0806) |
| N | 595 | 1820 | 2205 | 2730 |
| Rsquared | 0.870 | 0.808 | 0.893 | 0.866 |
| Quarter FE | Yes | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2.A.5: Cycle hire journeys ending near CSH

| | CSH | 20-200m | 200-400m | 200-600m |
|---|---|---|---|---|
| j<=-4 | 0.00315 | -0.0741 | 0.0325 | -0.0306 |
| | (0.129) | (0.0880) | (0.0798) | (0.0622) |
| j=-3 | 0.0447 | -0.125** | -0.0252 | -0.0499 |
| | (0.0922) | (0.0555) | (0.0581) | (0.0394) |
| j=-2 | 0.0439 | -0.0645 | 0.00682 | -0.0119 |
| | (0.0682) | (0.0414) | (0.0410) | (0.0216) |
| j=0 | 0.201 | 0.0174 | -0.00258 | -0.000451 |
| | (0.120) | (0.0503) | (0.0294) | (0.0212) |
| j=1 | 0.307 | 0.0453 | -0.0226 | -0.00850 |
| | (0.184) | (0.101) | (0.0513) | (0.0369) |
| j=2 | 0.465* | 0.126 | -0.0256 | -0.0197 |
| | (0.254) | (0.130) | (0.0763) | (0.0563) |
| j>=3 | 0.557 | 0.257 | 0.0174 | -0.0164 |
| | (0.376) | (0.190) | (0.106) | (0.0824) |
| N | 595 | 1820 | 2205 | 2730 |
| Rsquared | 0.900 | 0.835 | 0.911 | 0.883 |
| Quarter FE | Yes | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2.A.6: Cycling flow after a new cycle hire station opening

| | Ln Total Cycle |
|---|---|
| New cycle hire station | -0.0142 |
| | (-0.33) |
| N | 340 |
| Rsquared | 0.925 |
| Year FE | Yes |
| Site FE | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Figure 2.A.5: Opening of cycle hire stations before and after the construction of CSH

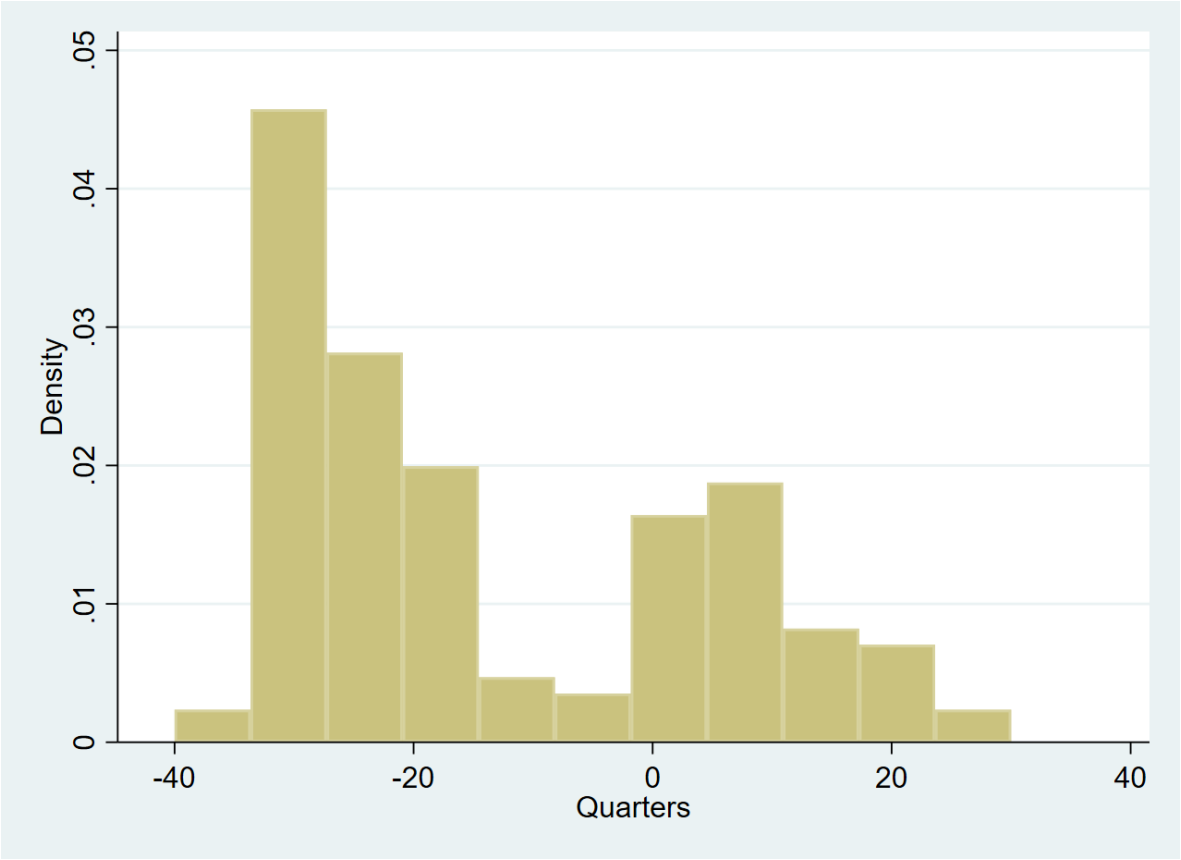


Table 2.A.7: Car displacement near CSH

| | CSH | 20-200m | 200-400m | 400-600m |
|---|---|---|---|---|
| j<=-4 | -0.242** | -0.136** | -0.0535 | 0.000599 |
| | (0.0621) | (0.0422) | (0.0544) | (0.0847) |
| j=-3 | -0.0788 | -0.0631*** | -0.0247 | -0.00462 |
| | (0.0430) | (0.0149) | (0.0138) | (0.0413) |
| j=-2 | -0.0900 | -0.0439** | -0.0236 | 0.00201 |
| | (0.0544) | (0.0119) | (0.0177) | (0.0451) |
| j=0 | -0.0183 | -0.0118 | -0.0204 | -0.0242 |
| | (0.0606) | (0.0276) | (0.0247) | (0.0351) |
| j=1 | 0.0146 | 0.0265 | -0.00286 | -0.0323 |
| | (0.0864) | (0.0531) | (0.0509) | (0.0758) |
| j=2 | 0.0554 | -0.00281 | -0.0508 | -0.0667 |
| | (0.166) | (0.117) | (0.0972) | (0.110) |
| j>=3 | 0.156 | 0.0557 | -0.0235 | -0.0878 |
| | (0.182) | (0.119) | (0.102) | (0.141) |
| N | 612 | 1496 | 2312 | 2890 |
| Rsquared | 0.896 | 0.915 | 0.915 | 0.919 |
| Year FE | Yes | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes | Yes |

SD clustered at year and cycle superhighway route level

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2.A.8: Bus displacement near CSH

| | CSH | <200m | <400m | <600m |
|---|---|---|---|---|
| j<=-4 | -0.182 | -0.216 | -0.238 | -0.255 |
| | (0.197) | (0.193) | (0.279) | (0.225) |
| j=-3 | -0.0867 | -0.0712 | -0.0894 | -0.120 |
| | (0.0730) | (0.0764) | (0.116) | (0.0906) |
| j=-2 | -0.0615 | -0.0695 | -0.0784 | -0.0888 |
| | (0.0539) | (0.0564) | (0.0910) | (0.0722) |
| j=0 | -0.0448 | 0.0596 | 0.105 | 0.0679 |
| | (0.0456) | (0.0630) | (0.0779) | (0.0378) |
| j=1 | -0.0326 | 0.128 | 0.173 | 0.140 |
| | (0.102) | (0.118) | (0.156) | (0.129) |
| j=2 | -0.0248 | 0.175 | 0.215 | 0.171 |
| | (0.146) | (0.159) | (0.189) | (0.128) |
| j>=3 | -0.0504 | 0.238 | 0.286 | 0.273 |
| | (0.161) | (0.206) | (0.273) | (0.224) |
| N | 604 | 1488 | 2301 | 2876 |
| Rsquared | 0.882 | 0.909 | 0.885 | 0.896 |
| Year FE | Yes | Yes | Yes | Yes |
| Site FE | Yes | Yes | Yes | Yes |

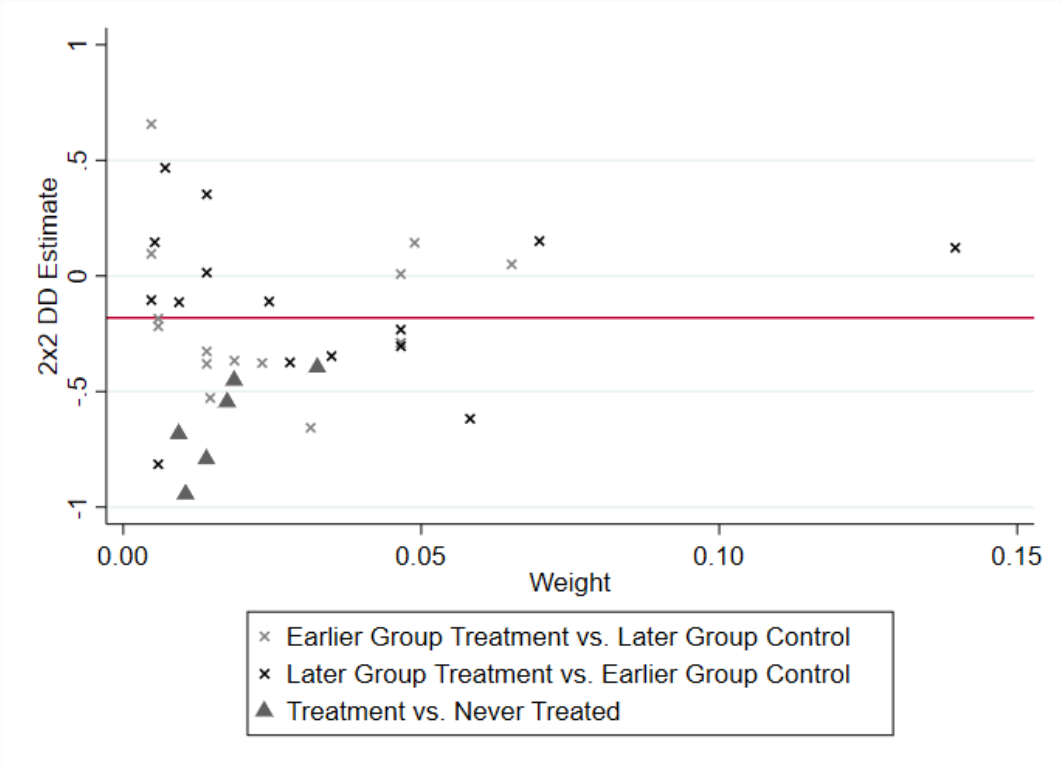
SD clustered at year and cycle superhighway route level

$^{*}$ $p < .1$, $^{**}$ $p < .05$, $^{***}$ $p < .01$

### 2.A.5 Accidents reduction

I further use the Goodman-Bacon decomposition to analyse the different components of the DiD estimator for bike accidents divided by cycling flow after a CSH opening. I find an overall ATT of -0.08, which is decomposed on an Earlier Treatment vs Later Control of -0.079, Later Treatment vs Earlier Control of 0.015 and Treatment vs Never treated of -0.573. The respective weights are 0.332, 0.559 and 0.109. All estimates (except from the problematic Later Treatment vs Earlier Control) are negative, which indicates that the cycling lanes are quite effective in reducing accidents and the effect is increasing over time.

Figure 2.A.6: Decomposition of DiD estimate for bike accidents



# 3 The impact of floods on local elections in England

## 3.1 Introduction

Flood risk is a serious global issue and particularly important for the United Kingdom (UK). It is estimated that around 5.4 million houses in England are at risk of flooding from rivers and the sea, surface water, or both, which represent around 20% of the total housing stock in the country. Annual flood damages in England are estimated to be around £1.1 billion (Priestley et al. 2017; Beltrán, Maddison, and Elliott 2019), and are expected to increase due to the greater intensity and frequency of floods induced by climate change. Evidence suggests that natural disasters, such as floods, can be framed and politicised in the media creating or exacerbating pre-existing tensions between political actors (Albrecht 2021). This means that natural disasters can become political events, affecting public opinions, policies, and political preferences.

Following the disastrous floods of 2007, described as a "national emergency", with rescue efforts as "the biggest in peacetime Britain", a key change in flood risk management has occurred in the UK formalized in the Flood and Water Management Act of 2010 (The Act hereafter). The Act, by creating the role of Lead Local Flood Authorities (LLFAs), gave new responsibilities to local authorities for managing flood risk and post-flood recoveries in local government areas. Becoming a highly politicized issue, the summer 2007 floods not only pressured politicians and policymakers but more likely have enabled voters to understand and differentiate between the responsibilities of different levels of government in the flood management arena. We thus use this context to study the impact of floods on retrospective accountability in a multi-level flood management system.

We construct a panel data of local election results from 2002 to 2018 at the electoral ward level. Wards are the smallest electoral unit in England. As of 2021, there are about 8,700 wards in total, representing on average 5,500 people (ONS, 2021). The ward election data comes from the LEAP - Local Elections Archive Project (Teale 2020). Voters in each ward elect one to three councillors to represent them in the local council - the governing body of a local authority. There are about 23 wards per local authority. Before the creation of the

LEAP dataset, local election results were only available at the local authority level. Measuring results at the ward level means that we can efficiently capture votes for non-dominant parties as smaller parties typically do not present a candidate in each ward. For our analysis, we merged these local electoral results with historic flood records from the Environment Agency (EA). Observing results at the ward level allows us to measure flood risk exposure much more precisely than using a district or regional level granularity. Finally, we also use the local authorities' budget from the Department for Levelling Up, Housing and Communities (DLUHC) to study whether and how public spending changes following floods. We capture policy change in flood risk management in our empirical analysis, by distinguishing two time periods before and after the introduction of the 2010 Management Act.

The 2010 change in flood management responsibilities, however, occurred contemporaneously with significant shifts in UK politics: the Labour party, after almost ten years in power, lost the 2010 General Election, which led to a Conservative-Liberal Democrats coalition and then a Conservative government in 2015. The Coalition government initiated an austerity program with a combination of public spending reductions and tax increases. Budget cuts on flood defences and their implications have received wide media coverage, especially in the aftermath of the 2015 floods (Albrecht 2021). The austerity programme has also been linked with a sharp cut in public spending in the poorest part of England and a rise of support for populist party such as the UK Independence party (UKIP) (Fetzer 2019).

We begin our analysis with a fixed effects panel model with a ward as unit of analysis, investigating whether floods affect local electoral outcomes. We control for time-invariant ward characteristics and national trends by adding ward fixed effects and year fixed effects. Our main outcome of interest is the share of votes for the incumbent party – the party that received the most votes in a previous local election. We find that the incumbent's vote share declines by about 0.7 percentage points if a flood occurs within the election year. The effects are more substantial for larger and longer floods in terms of area and duration, with 2.5-2.9 percentage points reduction. We also find that the incumbent's party affiliation with either the national government or local authority matters, as voters punish incumbents by up to 6 percent points if the incumbent shares the same party as national government. Voters punish

incumbents aligned with local authority only in the post-2010 period.

So, to disentangle the effects of the 2010 flood management reform from the effects of changes in the government and to shed light on the mechanisms, we utilize multiple empirical strategies. First, we extend our baseline analysis by analyzing candidates' shares of the votes by party: Conservatives, Labour, Liberal Democrats, UKIP, Green and others. We find that while candidates from major parties, especially when their party is in control of the local or national government, lose shares of votes after a flood. On the other hand, candidates from smaller parties, especially UKIP, gain shares of votes when the incumbent is aligned with the local and national government. This, in conjecture with the results for the Green party which loses shares of votes if the incumbent is aligned with the local council, would suggest that increased support for the UKIP has nothing to do with the environmental policy agenda.

Second, we exploit the set-up structures of local authorities, which could be either single-tier or two-tier councils - meaning that local authorities are either a single entity or divided into two entities with different flood responsibilities after 2010. Single-tier authorities or the upper level of two-tier councils become Lead Local Flood Authorities (LLFAs) responsible for the implementation of local flood risk management and leading role in emergency planning and recovery after floods, while district councils (lower level) become responsible for planning activities. We conjecture that if voters understand a shift in flood management responsibilities to local councils, then we should see differential effects of floods on LLFAs compared with district councils. We find no difference in these effects, suggesting that (i) either voters do not understand the nuances of the changes in the flood management policy; or (ii) they punish incumbents for other reasons, or both.

Third, we focus on contested elections. Local elections in England tend to be less contested in general, with an average turnout of 35% against 67% in national elections, with voters prone to vote for non-dominant parties (14% votes against 10% during national elections). More than half of the wards in our analysis do not change winning party during the time of our analysis. The seats in local governments are thus often coined as "safe seats", while local elections as "second-order" elections (Reif and Schmitt 1980). This suggests that only in

closely contested elections voters could exert accountability on local politicians if they wish of doing so. By focusing on contested elections, we find no difference between the effects of floods and party alignment on the incumbent's vote share in regular versus close elections. This points out to protest vote: voters express their dissatisfaction with the status quo, by casting vote for a party which has essentially no control over the local council.

So, we conclude that the incumbent's punishment is associated with votes for "protest parties" such as UKIP, crediting the hypothesis that natural disasters can bolster populist movements. Our findings are consistent with a rise in protest voting following the fiscal austerity reforms of 2010 (Fetzer 2019). Our results differ from recent findings by Cavalcanti (2018) for the case of Brazil which suggests that, after a period of drought, voters demand clientelism, i.e. they increase the vote share for local incumbent parties politically aligned with the central government to facilitate the flow of resources for relief and recovery. Instead and similarly to Lockwood, Porcelli, and Rockey (2022), our results suggest that constraints on local government fiscal policy in England mean that political control of local councils does not affect total expenditure – hence giving little incentives for rewarding incumbents politically aligned with local or national governments.

The rest of the paper is organized as follows. Section 3.2 discusses the related literature. Section 3.3 provides context, by discussing the structure of local government and flood risk management responsibilities across various government layers. Section 3.4 discusses the main data and identification strategy. Section 3.6 discusses the results, and Section 3.7 discusses the results on public expenditure and floods. Section 3.8 concludes.

## 3.2 Literature review

This paper is related to the extensive empirical literature which examines how voters evaluate political performance and respond to different types of information when deciding to re-elect a politician or a party. Voters' recall of politician performances is crucial to welfare-improving policies (Fiorina 1978). Voters use past policy outcomes to form expectations about their future welfare. The key question of interest is thus whether voters punish incumbent politicians for random events outside of their control ("blind retrospection"), punish based on the quality

of their responses ("attentive electorate"), or even "rally-around-the-flag" by increasing support to the party in control in times of crisis.

Achen and Bartels (2004) argue that voters also engage in blind retrospection - holding politicians responsible for events out of their control - including floods. However, flood events are not totally out of the control of policymakers. Studies show that voters tend to react to relief aid and other forms of more targeted and easily observable public spending, which also attract more media attention. For instance, Healy and Malhotra (2009) study how electors incentivize incumbents to invest in disaster preparedness and relief policies in the United States and find that voters only reward disaster relief spending. Masiero and Santarossa (2020) analyse the outcomes of municipal elections in Italy after earthquakes and find a massive advantage for the incumbents driven mainly by the ability to deliver relief and attract media coverage. In the same spirit, Besley and Burgess (2002) find that politicians are more responsive to disasters in areas with more newspaper circulation in India.

Although large relief spending tends to reward politicians, the important question is how long such gains in electoral support persist. Most studies have highlighted the myopia of voters. For instance, Cole, Healy, and Werker (2012), using rainfall, public relief, and election data from India, show that voters only respond to rainfall and government relief efforts during the year immediately preceding the election. In contrast, Bechtel and Hainmueller (2011), by exploiting the 2002 Elbe flooding in Germany, find that voters' rewards could last longer: the 25% increase in the vote share of the incumbent party in the 2002 election in affected areas persisted until the 2005 election, but completely vanished by the 2009 election.

The design of retrospective accountability, however, hinges upon the important assumption that voters can assign responsibilities to different levels of government and internalize those distinctions into their voting decisions. Previous research has shown that highly-politicised issues create information-rich environments that enable voters to differentiate between the responsibilities of different levels of government and influence their voting behaviour Wilson and Hobolt (n.d.).

Other papers have looked at the concordance between local incumbents and national government parties. Cavalcanti (2018) shows that following a natural disaster in Brazil, electors tend to reward incumbents politically aligned with the national government to ensure better disaster relief. In this context, the occurrence of a natural disaster creates a demand for clientelism where governments are also incentivised to prefer their own constituencies when providing disaster relief funds. We do not find such results in England, developed economy set-up. On the contrary, electors tend to punish the party in power. Our results are in line with a recent study by Lockwood, Porcelli, and Rockey (2022) which find little impact of the party control on local fiscal policy in England and Wales over the period 1998 to 2016. In general, as the party in control of the local councils has less discretionary power to increase spending than in other decentralised countries, there is little scope for clientelism.

Finally, this study also relates to the impact of natural disasters on political systems. Kaufmann et al. (2016) show that floods can lead to long-term institutional changes. Natural disasters can also weaken political systems and lead to more autocratic regimes (Rahman et al. 2017). In the context of the UK, large floods put pressure on the national government (Albrecht 2021).

## 3.3 Background

### 3.3.1 Elections and governance

We are interested in understanding if voters hold politicians accountable for floods in England. However, there are many actors are involved in flood management.

In England, local governments are represented by local authorities. The most common are local councils. They are composed of councillors elected every four years. Each councillor is representing a ward. Electors in each ward elect one or more councillors to represent them. Local elections use plurality voting – meaning that the candidates with the most votes win the seat. Electors cast votes for as many seats as there are being contested. Several wards form a council. The leader of the council is chosen from the political party with the most councillors. In 2022, there are 333 local authorities in England.

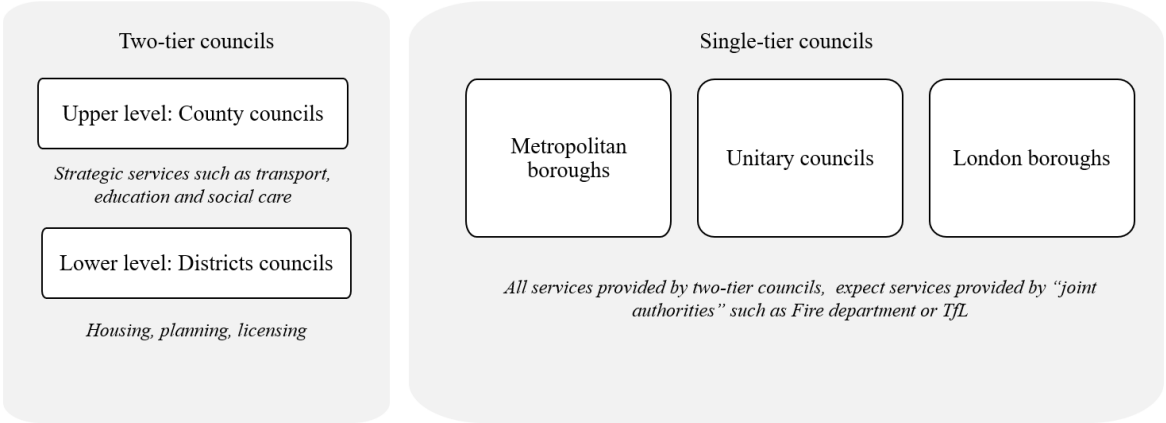
### 3.3.2 Structure and functions

There are two different forms of council set-ups as shown in Figure 3.3.1: single-tier councils and two-tier councils. Single-tier councils are comprised of one council, which carries out all local government functions and could be one of the following three types: metropolitan boroughs (there are 36 of them as of 2022), unitary councils (56) and London boroughs (32).

Two-tier councils are comprised of an upper level, the county, and a lower level, the district, with responsibilities being divided between two levels. County councils are responsible for strategic services such as schools, social services, public transportation, highways, fire and public safety, libraries, waste management and trading standards. District councils are responsible for more place-based services such as rubbish collection, recycling, council tax collection, housing, and planning applications. There are currently 33 country councils and 201 district councils. Within any area covered by one county, there will be approximately 5-7 district councils.

The unitary authorities provide all services mentioned above. In large metropolitan areas, some services, like fire, police and public transport, are provided through 'joint authorities'.

Figure 3.3.1: Local government structure and responsabilities



### 3.3.3 Managing flood risk: roles and responsibilities

The flood management responsibilities are shared at different levels of governance as shown in Figure 3.3.2. The Flood and Water Management Act 2010, the primary legislation in England relating to flood risk management (FRM) was introduced on 8 April 2010. It was intended to implement Sir Michael Pitt's recommendations following the widespread flooding of 2007 when more than 55,000 homes and businesses were flooded with insurance costs expected to be more than £3 billion (Pitt, 2008). The Act requires better management of flood risk and, most importantly, represents a key shift towards a localised flood risk management (FRM) agenda. National policies are delivered at the local level by Risk Assessment Management Authorities (RMAs).

Among these RMAs, counties (the upper level of two-tier authorities) and single-tier authorities are Lead Local Flood Authorities (LLFAs). LLFAs are responsible for developing and implementing local strategies for FRM, with responsibility for watercourses other than main rivers, surface water and groundwater. They also play a lead role in emergency planning and recovery after a flood event. LLFAs must act consistently with the national flood, coastal and environmental management strategy developed by the Environment Agency (EA) – a non-departmental public body.

District councils – the lower level of two-tier authorities are also Risk Assessment Management Authorities (RMAs). They are responsible for carrying out flood risk management works on minor watercourses and coordinating with LLFAs and other Risk Management Authorities.

Other non-elected RMAs are the Environment Agency, flood authorities, water and sewerage companies, internal drainage boards and highway authorities. Among those, the Environment Agency (EA) is the dominant actor and implements a strategic overview of FRM in England for all types of flooding and plays a key role in the distribution of national funding for defence and mitigation works. Both the EA and LLFAs are engaged in activities to raise community awareness and encourage the uptake of property-level resistance and resilience measures (Alexander et al., 2016).

At the national government level, The Department for Environment, Food and Rural Affairs (DEFRA) is responsible for creating policies for floods and coastal erosion, the Cabinet Office is in charge of emergency response planning, and the Department for Communities and Local Government deals with land-use and policy planning. Finally, the treasury is involved in the budget decision.

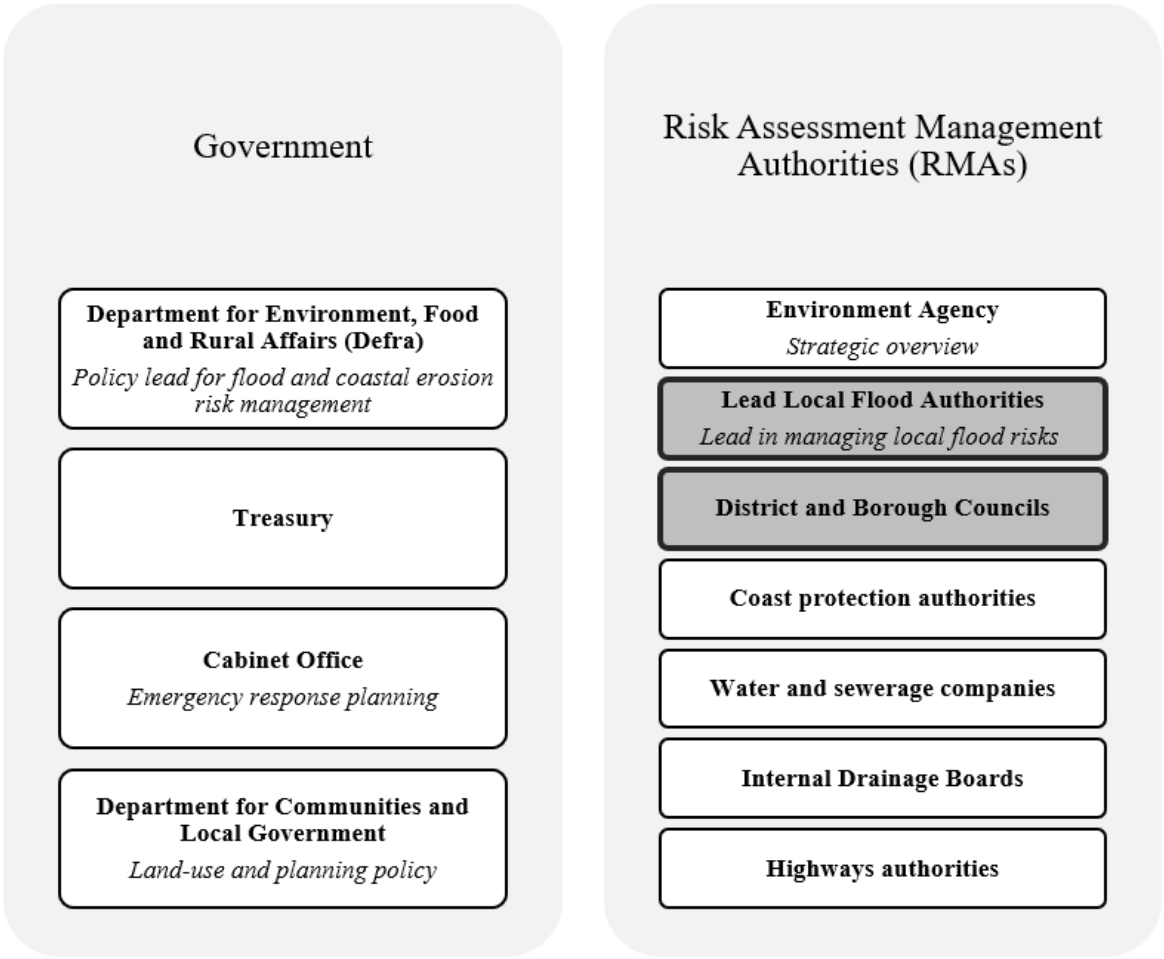
The key flood risk management strategies employed in FRM in England include prevention, defence and mitigation, preparation and response and recovery. The first two strategies aim to minimise the likelihood of flooding and people's exposure to flooding, while the latter strategies aim to minimise the consequences of flooding. It is worth stressing that local authorities play an essential support role in the community post-recovery period.

While the 2010 flood management established an ambitious plan to level flood defences, its application coincided with the austerity period. As a result, the national government's spending on floods and coastal erosion stayed at about £800 million per year between 2010 and 2018 (See Appendix Figure 3.A.3, Tily (2020)). In general, DEFRA's day-to-day budget was decreased and investment spending remained stable (Appendix Figure 3.A.4, Department for Environment (2022)).

At the local authority level, the Ministry of Housing, Communities & Local Government (MHCLG) data show an increase in spending of 59% for coastal erosion and 176% for flood defences for the 2010-2018 period (from about £50 to £140 million in flood defences and £55 to £85 million in cost protection, see Figure 3.A.5 and 3.A.6) (Clugston 2021). While flood spending has increased at the local level, revenues stayed stable or decreased, putting a strain on local authorities' finance. This is because the central government grants only cover a fraction of that spending, and grants have been decreased over the period. At the same time as the flood management act, the 2011 Localism Act lead to a decrease in local authority revenues by 16%. It was primarily driven by a decline in government grants by 37% (Figure 3.A.7, Atkins and Hoddinott (2022)). Councils can also apply for funds from the Environmental Agency, which awards the money on behalf of DEFRA. However, the bidding system has been criticised as more favourable to larger and more urban councils.

The responsibility of local authorities does not stop at the maintenance and construction of flood defences. They are the primary provider of social support at the local level. Floods can make households unable to live in their homes for months and have disastrous impacts on local businesses. The austerity-induced welfare reform of 2010 decreased the spending power of local authorities in deprived areas the most and led to an increase in votes for UKIP and the Brexit referendum (Fetzer 2019).

Figure 3.3.2: Flood management role and responsibilities




## 3.4 Data

To conduct this analysis, we use the results of the local elections compiled by the Local Elections ArchiveProject (LEAP) for 2002-2018. The database contains information about

all votes being cast, the total for each candidate, the name of the candidates and whether or not they were elected for each local election in England and Wales. Each election result is reported at the ward level. We merge the results to the corresponding ward shapefile for the period. There are 58,277 observations describing the number of votes and candidates for each ward and election. We discard observations without votes/candidates elected (1734) and exclude Wales as we do not observe all the elections.

We aggregate the number of votes at each election for four main categories: Conservatives (C), Labour (Labour), Liberal Democrats (LD), and Others, which enables us to compute the share of votes cast for each party. Where appropriate for our analysis, we will also decompose the Others category into its sub-categories, namely Green party, UKIP and Others (excluding Green and UKIP). The UK is often called a two (and a half) party system. It is based on the dominance of the two main parties: Conservative and Labour, and the consistent results of the Liberal Democrats party. These three parties control most councils in the UK. We then merge each ward election with a ward shapefile for the same year provided by the ONS.

We then create the intersection of the 2002 wards (base unit) and wards with elections for each subsequent year using the electoral wards shapefiles from the ONS geography. Then, for each year, we use how much of the ward area falls into the 2002 wards to allocate the correct share of votes to the 2002 wards. For example, if a ward boundary has not changed, then 100% of its area will fall into 2002 wards and all of its votes would be counted in the 2002 original wards. However, if a new ward was created in 2003 with 20% of its area in a 2002 ward and 80% in another 2002 ward, we allocate the 2003 votes to the 2002 wards using the share of 20% and 80%, respectively. We can thus re-aggregate all the results at the 2002 ward level even if wards have been divided, merged or even completely redesigned. Our final dataset includes about 8,000 wards over 17 years of local elections. We show an example of wards that have changed in most urban areas in Appendix Figure 3.A.1.

In the final panel, we use total votes to create shares of votes for each party. Alignment with the local government is defined as wards whose incumbent party (the party that received the most votes at the last election) is from the same majority as the local council majority. For

example, if the council is led by the Conservative – meaning that the majority of councillors across the wards that compose the local council are Conservative, then a ward is aligned with the council if their incumbent party is Conservative. Alignment with the national government is defined as wards whose incumbent party is Labour before 2010, Conservative and Liberal Democrats up to 2015 and Conservative after 2015.

Next, we overlay the ward shapefiles with GIS data showing Recorded Outlines of Individual Flood Events (ROIFE) in England from the Environment Agency (2022b) (recording started in 1946). We show in Appendix Figure 3.A.2 that 60% of the wards are flooded at least once during the 1992-2018 period. We look at floods up to 10 years before the start of our dataset to construct a full history and investigate how long the electorate's recall is. We show later in the analysis that only wards within a year of a local election matter. In our sample, 3% of the wards holding elections have been flooded within a year of a local election. We also calculated intensity measures such as the duration of the floods in days, the share of the wards flooded and the total area of the floods. We find that large floods drive our results in terms of duration and size.

In addition to flood data, we also overlaid the electoral wards with the Spatial Defences from the Environment Agency (2022a), which contains all the structures managed by the environmental agency, their condition and their building date. Unfortunately, most records have been added after implementing the Flood management reform in 2010/2012 and flattened afterwards, which makes them difficult to use in a panel analysis. We also calculated the maximum risk of Flooding from Rivers and Sea from the Environment Agency (2022c). We classified each ward at risk as very low, low, medium, or high based on flood return rates. Very low, low, medium and high risk means that each year, there is a chance of flooding of less than 0.1%, 1%, 3.3% and above 3.3% respectively. Each ward receives the rating of the maximum risk level is exposed. For example, if a ward has two flooding zones, one in low and the other one in medium, it is given a medium risk rating.

Finally, to complement the analysis at the electoral ward level, we constructed a panel data of spending from the Department for Levelling Up and Communities (2022) at the local

authority level that we linked with elections and flood history.

### 3.5 Identification strategy

Our empirical investigation focuses on local election results at the ward level. During elections, councillors are elected for 4-year terms using the first past the post system. Councillors represent divisions/ wards. Elections to councils are held on the first Thursday in May. A specificity of England is that wards have different voting frequencies. The local government can be elected in one of the three following ways: elect all the local councillors every 4 years, elect half the local councillors every 2 years or elect one-third of the local councillors every year for 3 years and hold no elections in the 4th year. We thus obtain an unbalanced panel of local election results at the ward level from 2002 to 2018.

Our main outcome is the share of votes for the incumbent party – meaning the party that got the most votes in a ward during the last election. We control for time-invariant ward characteristics and national trends by adding ward fixed effects and year fixed effects. Our main dependent variable is a dummy for floods within one year of the election. We also use the count of floods in the ward, but as most wards get only flooded once on average, the results are fairly similar. We then add flood characteristics such as the duration and size of the flooded area.

One of the main concerns would be if wards in flooded areas had significantly different voting behaviours. We show in Table 3.5.3 that there are also no significant differences in terms of political composition.

We also restrict our sample to wards that get flooded within a year of a local election between 2002-2018, so that we compare similar wards in our regressions.

In addition, in Table 3.5.1, we compare our regression sample with all other wards in England using 2001 census variables. We find that our flooded wards have slightly more households living in terraced houses (27 versus 24%), and they have a larger share of the population that identified themselves as white. However, there are no significant differences

in education achievement, occupation and the share of urban population. They are not particularly UKIP strongholds. We also look at the long difference between 2001 and 2011 in Table 3.5.2. Again, there are no major differential trends from the rest of the population.

Wards with high flood risk are more likely to end up in our sample: over 20% of treated wards are marked as being in a flood zone against 14% for the rest of the country. Only 2% of the wards with a very low flood risk get flooded within a year of the elections, while 5% of wards with a high flood risk get flooded before an election. The number of flood defences (each flood defence is defined as a single asset by the environmental agency) and the length of the flood defences also increase with flood risks as shown in Table 3.5.4. Overall, in Table 3.5.4, we can observe that riskier areas are getting flooded more often, but they are also allocated more resources. Therefore, without accounting for flood defences, we are capturing a lower bound coefficient of the impact of floods on local election results.

Table 3.5.1: Balance table using 2001 census

| | Flooded before an election (N=1161) | | All others (N=7028) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Diff |
| Age | 39.56 | 3.14 | 39.64 | 3.64 | -0.074 |
| Wholehouse | 28.85 | 18.87 | 31.57 | 20.77 | -2.723 |
| Semidetached | 33.68 | 11.97 | 32.82 | 13.01 | 0.864 |
| Terraced | 27.14 | 15.08 | 23.77 | 14.88 | 3.362*** |
| Flat | 10.36 | 9.13 | 10.89 | 11.59 | -0.532 |
| Converted | 2.93 | 4.03 | 3.54 | 6.72 | -0.617* |
| Commercial | 1.12 | 0.93 | 1.11 | 1.04 | 0.014 |
| Caravan | 0.78 | 1.85 | 0.61 | 1.58 | 0.172* |
| No qualification | 27.97 | 9.15 | 28.26 | 8.39 | -0.292 |
| Level1 | 16.55 | 3.18 | 16.94 | 3.36 | -0.386* |
| Level2 | 20.09 | 2.89 | 20.08 | 2.89 | 0.007 |
| Level3 | 8.17 | 4.38 | 7.91 | 3.22 | 0.263 |
| Level4 | 20.21 | 9.08 | 19.65 | 9.07 | 0.560 |
| Other qualification | 7.01 | 1.38 | 7.16 | 1.43 | -0.148 |
| White | 95.69 | 7.15 | 94.21 | 10.93 | 1.482** |
| Mixed | 0.84 | 0.66 | 1.00 | 0.96 | -0.159** |
| Asian | 2.37 | 5.77 | 2.74 | 7.20 | -0.370 |
| Black | 0.53 | 1.25 | 1.40 | 4.12 | -0.863*** |
| Other ethnicity | 0.57 | 0.76 | 0.66 | 0.97 | -0.091 |
| Agriculture | 2.25 | 2.97 | 2.54 | 3.27 | -0.288 |
| Mining | 0.36 | 0.64 | 0.29 | 0.47 | 0.073 |
| Manufacturing | 15.29 | 5.69 | 15.03 | 5.65 | 0.259 |
| Electricity | 0.77 | 0.54 | 0.70 | 0.51 | 0.075** |
| Construction | 6.89 | 1.68 | 7.09 | 1.85 | -0.199 |
| Trade | 16.95 | 3.05 | 16.73 | 2.98 | 0.214 |
| Hotels | 5.22 | 2.97 | 4.92 | 2.32 | 0.292 |
| Transport | 6.70 | 2.72 | 6.71 | 2.40 | -0.010 |
| Finance | 3.75 | 2.04 | 4.27 | 2.76 | -0.513** |
| Realestate | 12.71 | 5.16 | 12.54 | 4.66 | 0.176 |
| Public | 5.59 | 2.87 | 5.70 | 3.82 | -0.113 |
| Education | 8.09 | 3.25 | 7.74 | 2.48 | 0.353 |
| Health | 10.50 | 2.43 | 10.69 | 2.44 | -0.194 |
| Other industry | 4.95 | 1.14 | 5.07 | 1.68 | -0.125 |
| Urban | 63.43 | 40.93 | 64.83 | 42.18 | -1.400 |

Standard errors clustered at the local authority level
The following variables have been used: average age, type of dwelling, educational qualification, industry, and urban share.
All variables are expressed as shares.

Table 3.5.2: Balance table using log change between census 2001 and 2011

| | Flooded before an election (N=1161) | | All others (N=7028) | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Diff |
| Δ Age | 0.03 | 0.04 | 0.03 | 0.04 | -0.002 |
| Δ Wholehouse | -0.05 | 0.18 | -0.03 | 0.18 | -0.010 |
| Δ Semidetached | -0.05 | 0.12 | -0.05 | 0.11 | -0.007 |
| Δ Terraced | -0.06 | 0.16 | -0.04 | 0.17 | -0.016** |
| Δ Flat | 0.17 | 0.36 | 0.17 | 0.41 | 0.004 |
| Δ Converted | -0.07 | 0.42 | -0.06 | 0.53 | -0.010 |
| Δ Commercial | -0.12 | 0.44 | -0.11 | 0.54 | -0.011 |
| Δ Caravan | -0.46 | 1.98 | -0.41 | 2.01 | -0.053 |
| Δ No qualification | -0.24 | 0.09 | -0.24 | 0.10 | -0.003 |
| Δ Level1 | -0.23 | 0.10 | -0.23 | 0.10 | -0.002 |
| Δ Level2 | -0.25 | 0.13 | -0.24 | 0.12 | -0.009 |
| Δ Level3 | 0.46 | 0.25 | 0.46 | 0.25 | 0.007 |
| Δ Level4 | 0.37 | 0.15 | 0.37 | 0.14 | -0.000 |
| Δ Other qualification | 0.22 | 0.23 | 0.22 | 0.28 | 0.000 |
| Δ White | -0.04 | 0.07 | -0.05 | 0.08 | 0.005 |
| Δ Mixed | 0.61 | 0.40 | 0.62 | 0.53 | -0.004 |
| Δ Asian | 1.14 | 0.99 | 1.15 | 1.04 | -0.018 |
| Δ Black | 0.86 | 1.45 | 0.77 | 1.38 | 0.092 |
| Δ Other ethnicity | -0.46 | 1.35 | -0.56 | 1.52 | 0.108 |
| Δ Agriculture | -1.01 | 0.84 | -1.04 | 1.04 | 0.028 |
| Δ Mining | -0.42 | 1.64 | -0.53 | 1.94 | 0.117* |
| Δ Manufacturing | -0.49 | 0.16 | -0.51 | 0.21 | 0.022 |
| Δ Electricity | 0.64 | 0.55 | 0.71 | 0.58 | -0.071** |
| Δ Construction | 0.13 | 0.14 | 0.14 | 0.15 | -0.004 |
| Δ Trade | -0.06 | 0.09 | -0.05 | 0.09 | -0.013*** |
| Δ Hotels | 0.11 | 0.16 | 0.12 | 0.18 | -0.010 |
| Δ Transport | 0.26 | 0.25 | 0.23 | 0.23 | 0.029 |
| Δ Finance | -0.11 | 0.23 | -0.12 | 0.24 | 0.012 |
| Δ Realestate | -0.48 | 0.20 | -0.49 | 0.21 | 0.001 |
| Δ Public | 0.70 | 0.25 | 0.70 | 0.26 | -0.002 |
| Δ Education | 0.25 | 0.13 | 0.26 | 0.14 | -0.011* |
| Δ Health | 0.15 | 0.11 | 0.14 | 0.12 | 0.003 |
| Δ Other industry | -0.01 | 0.15 | 0.00 | 0.17 | -0.014* |
| Δ Urban | 0.00 | 0.00 | 0.00 | 0.00 | 0.000 |

Standard errors clustered at the local authority level
The following variables have been used: average age, type of dwelling, educational qualification, industry, and urban share.
All variables are expressed as the difference in log share.

Table 3.5.3: Share votes by political party

| | Flooded before an election | All others |
|---|---|---|
| Share votes Conservative | 35.09 | 38.01 |
| | (19.48) | (20.66) |
| Share votes Labour | 29.76 | 28.00 |
| | (21.18) | (21.38) |
| Share votes Liberal Democrats | 18.98 | 18.57 |
| | (17.84) | (18.50) |
| Share votes UKIP | 4.882 | 4.241 |
| | (8.658) | (8.463) |
| Share votes Green | 3.376 | 3.212 |
| | (6.351) | (6.402) |
| Share votes Other | 7.917 | 7.971 |
| | (14.97) | (15.39) |
| Observations | 8856 | 45758 |

Table 3.5.4: Risk of flooding vs share of wards flooded and flood defences

| | Very Low | Low | Medium | High |
|---|---|---|---|---|
| Flooded | 0.0233 | 0.0329 | 0.0382 | 0.0494 |
| | (0.151) | (0.178) | (0.192) | (0.217) |
| Total assets | 34.71 | 31.85 | 38.72 | 45.90 |
| | (30.71) | (30.48) | (40.22) | (45.22) |
| Total length | 14.34 | 17.92 | 25.70 | 30.31 |
| | (11.65) | (19.68) | (23.03) | (26.63) |
| Observations | 344 | 1460 | 2512 | 3850 |

Very low, low, medium and high risk means respectively a 0.1%, 1%, 3.3% and greater than 3.3% chance of flooding each year. This takes into account the effect of any flood defences in the area. These defences reduce but do not completely stop the chance of flooding as they can be overtopped, or fail. Total assets is the number of flood defences currently owned, managed or inspected by the EA. They can be both man-made or natural defences. Length is measured in km.

## 3.6 Local elections results

### 3.6.1 Voters' recall

The first step of our analysis is identifying which floods impact local election results - if any. We explore two dimensions: the length of voters' recall and the severity of the flood. The reasoning behind this first step is that many single event studies have found a limited voters' recall after a flood - putting into question the rationality of voters' behaviours - and have also focused on single large events.

For the first part of the investigation, we do not limit the sample to wards flooded within a year of a local election, but to any ward flooded between 1992 and 2018 to study the length of voters' recall. Our basic specification is the following:

$$ShareIncumbent_{w,t} = \sum_{j}^{T} \alpha_j Flooded_{w,t+j} + \gamma_w + \delta_t + \epsilon_{wt}$$

Share incumbent is the share of votes for the incumbent party in ward $w$ and election $t$ – the incumbent is defined as the leading party from the previous election. $Flooded_{w,t+j}$ is a dummy equal to one if the ward was flooded within $t + j$ years of the election. We consider floods happening up to 2 years before an election and one year after. Our main coefficient of interest is $Flooded_t$, flooded within the year of a local election. $\gamma_w$ and $\delta_t$ are ward and year fixed effects.

We extend the model by restricting the sample using flood characteristics such as duration in days or the size of the flooded area. Table 3.6.1 shows the results for wards flooded within a year of the election (column 1), and then floods happening a year after and up to two years before an election (column 2). Next, we restrict the analysis to large floods: over the 75th percentile in terms of duration (column 3) and over the 75th percentile in terms of flooded area (column 4). Finally, we look at any other floods: under the 75th percentile in terms of duration (column 5) and under the 75th percentile in terms of flooded areas (column 6).

In concordance with the literature, we find evidence that voters only consider floods happening within a year of the elections. The effect of floods on the share of votes overall is relatively small (Columns 1 and 2 - 0.7 percentage points decrease in support). The effect is quadrupled when we restrict to large floods (columns 2 and 3), but disappears when we restrict to smaller events. There is no impact of floods happening after the elections (except column 3 – but it does not hold for other measures) or more than a year before the elections.

From now on, we focus on wards flooded within a year of an election to make sure that our treated group (flooded) is as similar as possible to our control group (not flooded but has been or will be flooded within a year of a local election). When looking closer at the characteristics of the floods in Appendix Table 3.A.4, we find that duration (column 1) and size (column 2) matter. The size of the area flooded seems to matter more than the duration of the floods (columns 3 and 4). In Appendix Table 3.A.4, we find a similar pattern using a dummy variable for floods in the 75th percentile of their distributions in days and areas.

These results indicate that voters' recall is limited to a year before the election. Larger and longer floods elicit a stronger response. The limited recall of electors cast some doubt on the perfect rationality of electors. As we cannot measure to which extent local and national governments could have prevented the floods, it is not possible to judge if punishing the incumbent party is a rational decision. However, the lack of recall shows they are not taking into account the full information about flood history but only the most recent events.

### 3.6.2 Flood management responsabilities

In the next step, we look at flood management responsibilities. As mentioned earlier, flood management responsibilities are divided between the local and the national government in England. We thus examine the interaction between the ward's incumbent party and the local council's majority before and after 2010. We also decompose the results for each party recorded in the dataset to show how other parties (non-incumbent parties) capture some of the votes after an election.

We start by looking at alignment with the local government before and after 2010. We run

Table 3.6.1: Large and long-lasting events on the share of vote for the incumbent

| | (1) No lags | (2) All | (3) Long events | (4) Large events | (5) Short events | (6) Small events |
|---|---|---|---|---|---|---|
| Flooded t+1 | | -0.195 | -1.606** | -1.067 | -0.0314 | -0.213 |
| | | (0.306) | (0.731) | (0.741) | (0.436) | (0.437) |
| Flooded t | -0.703* | -0.692* | -2.919*** | -2.491*** | 0.0414 | -0.173 |
| | (0.382) | (0.397) | (0.795) | (0.902) | (0.432) | (0.424) |
| Flooded t-1 | | 0.149 | 0.0394 | 0.225 | 0.0982 | 0.110 |
| | | (0.388) | (0.681) | (0.840) | (0.457) | (0.420) |
| Flooded t-2 | | 0.0642 | 0.261 | -0.217 | -0.0249 | 0.116 |
| | | (0.343) | (0.690) | (0.935) | (0.374) | (0.352) |
| N | 18512 | 18512 | 18512 | 18512 | 18512 | 18512 |
| Rsquared | 0.0583 | 0.0583 | 0.0593 | 0.0588 | 0.0581 | 0.0581 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
Sample is any wards flooded between 1992-2018
Year t is the year before a local election. Column 1 only includes floods hapenning a year before a local election.
Column 2 includes any floods. Column 3 and 4 restrict to floods above the 75th percentile in terms of duration and area flooded respectively.
Column 5 and 6 restrict to floods above the 75th percentile in terms of duration and area flooded.

the following regression for all years, and then pre and post-2010 separately:

$$ShareIncumbent_{w,t} = \beta_1 Flooded_{w,t} + \beta_2 Govenment_{w,t} + \beta_3 Flooded_{w,t} \times Government_{w,t} + \gamma_w + \delta_t + \epsilon_{wt} \quad (1)$$

$Shareincumbent_{w,t}$ is the share of votes for the incumbent party in ward $w$ and election $t$ – the incumbent party is defined as the leading party from the previous election. $Flooded_{w,t}$ is a dummy for flooded within a year of a local election. $Govenment_{w,t}$ is a dummy for alignment of the incumbent party with either the majority of the local council or the party in power at the national government level.

Table 3.6.2 reports the results of the estimation equation (1). Compared with the baseline results presented above, these results show that voters punish incumbents even more so if they are affiliated with the national government, with incumbents losing between 1.8 and 6.3 percentage points in votes after a flood. Moreover, voters punish incumbents in both pre- and

post-2010 that is irrespective of which party were leading the government in the UK. The national government has always been responsible for creating policies for flood and coastal erosion risk management in England and remains to be a major funder of emergency and flood prevention policies. Therefore, the occurrence of floods is perceived as a failure of the national government by voters. We, however, note that the magnitude and significance of the punishment effect on incumbents aligned with the national government decreases for the post-2010 period.

Table 3.6.3 is a counterpart of Table 3.6.2 and reports the results for the case when the incumbent has the same party affiliation as the local authority. Incumbents who share the party affiliation with the local council lose up to 1.9 percentage points (column 1), and up to 4 percentage points in post-2010 (column 3). Incumbents in wards aligned with the local government however do not lose any votes compared with non-aligned incumbents due to the occurrence of floods in the pre-2010 period (column 2). Both the decentralisation of flood management and the effect of the austerity policy post-2010 might be at play here. Voters could be holding the local government responsible for the failure of preventing the flood or for poor post-flood recovery initiatives. It is also possible that the series of public spending on local services, including flood mitigation and relief efforts, initiated by the coalition government in the post-2010 period made voters willing to punish the incumbent. In the next section, we shed light on the mechanisms which drive these results.

These findings are consistent with a recent study by Albrecht (2021): he explores the relationship between media framing of two floods in 2005 and 2015 in the UK. He finds that during the first period, the local and national governments were blamed partially for the disaster, but mostly for not doing enough. The opposition was harsher on the national government. After 2015, the overall tone was much more negative, with austerity and claims about EU spending being used to explain the poor state of flood defences. Media and opposition politicised the floods to fit the greater narrative of austerity and Brexit.

Table 3.6.2: Concordance with the national government - Labour vs. Conservative government

| | (1) | (2) | (3) |
|---|---|---|---|
| | All years | Pre-2010 | Post-2010 |
| Flooded | 0.0337 | 1.847*** | -0.814 |
| | (0.535) | (0.714) | (0.760) |
| NatGov | -5.652*** | -0.207 | 0.453 |
| | (0.311) | (0.906) | (0.846) |
| Flooded × NatGov | -3.117*** | -6.924*** | -1.864* |
| | (0.728) | (0.957) | (1.000) |
| N | 8856 | 4159 | 4697 |
| Rsquared | 0.136 | 0.0691 | 0.128 |
| Year FE | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election. LocalGov is a dummy for concordance of the ward's incumbent with the local council majority's party.

Table 3.6.3: Concordance with the local government - Labour vs. Conservative government

| | (1) | (2) | (3) |
|---|---|---|---|
| | All years | Pre-2010 | Post-2010 |
| Flooded | -0.297 | 0.172 | 0.324 |
| | (0.565) | (0.755) | (0.784) |
| LocGov | 3.976*** | 2.064*** | 3.432*** |
| | (0.408) | (0.553) | (0.564) |
| Flooded × LocGov | -1.883** | -1.097 | -3.962*** |
| | (0.741) | (1.046) | (0.931) |
| N | 8856 | 4159 | 4697 |
| Rsquared | 0.0883 | 0.0600 | 0.142 |
| Year FE | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election. LocalGov is a dummy for concordance of the ward's incumbent with the local council majority's party.

### 3.6.3 Protest voting

In this section, we explore three different empirical approaches, which all point out to one mechanism - protest voting, - behind our results on changes in incumbent vote shares in local elections following an occurrence of floods.

First, we examine the effect of flood occurrence on vote shares of incumbents affiliated with a specific party: Conservative, Labour, Liberal Democrat and Other in the pre- and post-2010 periods, by estimating the following model:

$$\begin{aligned} ShareIncumbent_{w,t} =& \beta_1 Flooded_{w,t} + \beta_2 LocalGov_{w,t} + \beta_3 NatGov_{w,t} + \beta_4 Post + \\ & \beta_5 Flooded_{w,t} \times LocalGov_{w,t} + \beta_6 Flooded_{w,t} \times LocalGov_{w,t} + \\ & \beta_7 Flooded \times Post + \beta_8 LocalGov \times Post + \beta_9 NatGov \times Post + \\ & \beta_{10} Flooded \times LocalGov \times Post + \beta_{11} Flooded \times NatGov \times Post + \\ & \gamma_w + \delta_t + \epsilon_{wt} \end{aligned} \tag{2}$$

where $ShareIncumbent_{w,t}$ is the vote share of the incumbent's party, $Flooded$ is a dummy which equals one, if a ward is flooded within a year of an election, $LocalGov$ is a dummy variable for incumbent's party alignment with local council's, $NatGov$ is a dummy variable for incumbent's party alignment with national government's, $Post$ is a dummy variable that equals one if elections occur in the post-2010 period. The variables of interest are: $\beta_5$, $\beta_6$, $\beta_{10}$ and $\beta_{11}$, where the former two measure the average effect of the occurrence of floods on the incumbent's share of votes under the case of party alignment with the local council and national government in pre-2010, respectively; the last two measure the additional differential effect compared to the pre-2010.

Table 3.6.4 reports the results of estimates of the interaction between flooded and alignment with the local and national government before 2010 $\beta_5$ and $\beta_6$; the same interactions after 2010 $\beta_{10}$ and $\beta_{11}$.

The results of the first column confirm our earlier findings that incumbents aligned with national government lose votes in both periods. This is because interaction term $Flooded \times NatGov$ is statistically significant, while interaction term $Flooded \times NatGov \times Post$ is not statistically significant, meaning that there is no significant differential effect in the share of votes of incumbents being aligned with the national government and flooded in the post-2010 period. The results on the effect of alignment with the local council is also confirmed: $Flooded \times LocGov$ is not statistically significant, while $Flooded \times LocGov \times Post$ is negative and statistically significant, implying that following a flood aligned with local councils incumbents are punished in the post-2010 period.

Moving to the results on the incumbent's vote share for individual parties (columns 2-5), we note that Conservatives incumbents lose votes after a flood, in both periods, and even more so when they held the national office (as a Coalition until 2015, and solely after 2015). Labour incumbents also lose votes in the pre-2010 when Labour held the national office, but in the post-2010, when they were no longer majority, they benefited from floods compared with the first period. In stark contrast to the major parties, other parties' incumbents in column 5 benefit from floods and gain votes in the post-2010.

By looking at the breakdown of votes within the Other category reported in Table 3.6.5, we see that UKIP has captured the decrease in votes from non-UKIP incumbents aligned with the local and national governments. The results from the Green party incumbents confirm that such an increase in votes for the UKIP party has nothing to do with anything in relation to climate change or environmental policy agenda. During that period, UKIP came from 4 councillors to 147 councillors in 2014 and won another 163 and 176 in 2014 and 2015, respectively. Other non-major parties lost their seats during the period. UKIP was successful in rallying places that have been most affected by austerity and promoting Brexit as a solution to budgetary cuts (Fetzer 2019). It is not surprising that UKIP managed to use floods to promote its platform successfully given the political framing: they blame the local and national governments for not doing enough and cutting flood management budgets. For example, after the 2014 floods, UKIP promised to increase spending in flood defences by cutting the overseas

aid budget[11].

While most of these results point out to an effect driven by protest voting, it is still possible that results after 2010 are driven by the flood management reform. If this is the case, electors should punish authorities that have more flood management responsibilities – counties and one-tier authorities (LLFAs). We, therefore, investigate if the effects of an occurrence of floods on the vote share are driven by the newly created LLFAs with results reported in Appendix Table 3.A.5. The coefficient estimates of $Flooded \times NatGov \times LLFAs$ in column 1 and $Flooded \times LocalGov \times LLFAs$ in column 2 are not statistically significant. Therefore, we haven't found differences between the authorities primarily responsible for local flood management - the LLFAs (one-tier authorities and county councils) and the authorities that cooperate with them (district councils). However, these results could be also driven by the fact that voters do not have a very clear understanding of the responsibilities division. Overall, the lack of significance renders the test inconclusive.

Another concern would be if more wards with a more secure affiliation with the local or the national government (safe seats) received more money for flood management and flood defences. It is linked to the concerns about the exogeneity of floods – while meteorological conditions are arguably exogenous to political settings, it is possible that wards affiliated with the party in power at the local and national government receive more funds to build flood defences. Wards with a less secure affiliation would be more likely to vote against the government in the aftermath of a flood and also less likely to have received flood defences in the past. It is also part of the protest story: if voters punish the incumbent in hopes of changing the local council's majority, they might be more likely to do so where the margins of victory are smaller.

In Appendix Table 3.A.8, we thus look at the difference between close elections (the margin of victory between the leading party and runner-up party is less than 10 percentage points) and find no difference between "safe seats" and close elections. It indicates that voters in safe seats are as likely to punish the incumbent as in close elections. Again, it points to protest

[11] https://www.theguardian.com/politics/2014/feb/09/nigel-farage-uk-aid-budget-somerset-flood-victims

voting: voters punish the incumbent and vote for a party that controls virtually no councils over the period.

In England, these results seem to point to a protest voting story. During local elections, incumbents aligned with the local or the national government lose votes after a natural disaster, and candidates from protest parties capture part of this discontentment. It would be interesting to look at the impact of floods on turnout. Unfortunately, there is no publicly available dataset of local election turnout at the ward level.

### 3.6.4 Floods history

In the next section, we look at heterogeneity in flood response by flood frequency. Electors in wards that get flooded more often might react less strongly to floods as they adapt to the risk in their area. In Appendix Table 3.A.7, we look at the interaction with flood frequency and find no significant difference between wards that get flooded more often and those that only get flooded less often.

## 3.7 Local authorities' expenditures results

We instead turn to flood defences and local authority expenditures to explain these results. A concern is that wards affiliated with the government might receive more funds to protect themselves against floods. A recent paper by Lockwood, Porcelli, and Rockey (2022) indicates that given the centralisation of the local government system and the limited ability of the local authorities to raise money, clientelism is less of a concern in the UK.

Unfortunately, the flood defences register maintained by the Environmental agency only shows a large increase in registered assets in 2012/2013 when it became mandatory to list assets in the database but has no new assets registered since (see Appendix Figure 3.A.8.

We then look at spending after a flood at the local authority level depending on the affiliation with the national government. We collect a dataset of expenditures at the local authority level for 343 councils $i$ for the year $j$ 2007 to 2017.

Table 3.6.4: Concordance with the local and national government pre and post 2010

| | (1) Incumbent | (2) Conservative | (3) Labour | (4) LD | (5) Other |
|---|---|---|---|---|---|
| Flooded | 2.073*** | 2.245*** | -1.649*** | -0.141 | -0.456 |
| | (0.766) | (0.544) | (0.446) | (0.646) | (0.754) |
| LocGov | 3.286*** | 1.026*** | 1.681*** | -0.998** | -1.708*** |
| | (0.513) | (0.390) | (0.360) | (0.437) | (0.539) |
| NatGov | -6.242*** | -1.198*** | 1.199** | 0.311 | -0.312 |
| | (0.642) | (0.445) | (0.487) | (0.557) | (0.641) |
| Flooded × LocGov | 0.978 | -0.139 | 0.702 | -0.688 | 0.124 |
| | (0.970) | (0.894) | (0.710) | (0.902) | (1.035) |
| Flooded × NatGov | -5.108*** | -0.382 | -1.387* | 0.132 | 1.637 |
| | (0.934) | (0.837) | (0.835) | (0.878) | (1.029) |
| Post | 0.0328 | -3.189*** | 9.551*** | -10.27*** | 3.904*** |
| | (0.730) | (0.434) | (0.449) | (0.613) | (0.626) |
| Flooded × Post | -0.820 | -2.202** | 1.302* | -0.0413 | 0.941 |
| | (1.466) | (0.950) | (0.743) | (0.947) | (1.233) |
| LocGov × Post | 0.892 | -1.652*** | -0.588 | -0.940* | 3.180*** |
| | (0.576) | (0.461) | (0.456) | (0.519) | (0.647) |
| NatGov × Post | 2.039* | 1.935*** | -3.397*** | 0.640 | 0.823 |
| | (1.174) | (0.725) | (0.728) | (0.954) | (1.041) |
| Flooded × LocGov × Post | -6.004*** | -2.304* | -1.764* | -0.436 | 4.503*** |
| | (1.470) | (1.201) | (0.944) | (1.165) | (1.549) |
| Flooded × NatGov × Post | 1.496 | -5.270*** | 2.056** | -0.0368 | 3.251** |
| | (1.465) | (1.186) | (1.037) | (1.142) | (1.507) |
| N | 8856 | 8856 | 8856 | 8856 | 8856 |
| Rsquared | 0.0852 | 0.0887 | 0.279 | 0.303 | 0.100 |
| Year FE | No | No | No | No | No |
| Ward FE | Yes | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election.
LocGov is a dummy for concordance of the ward's incumbent with the local council majority's party.
NationalGov is a dummy for concordance of the ward's incumbent with the national government's party.
Post is a dummy for elections happening after 2010.

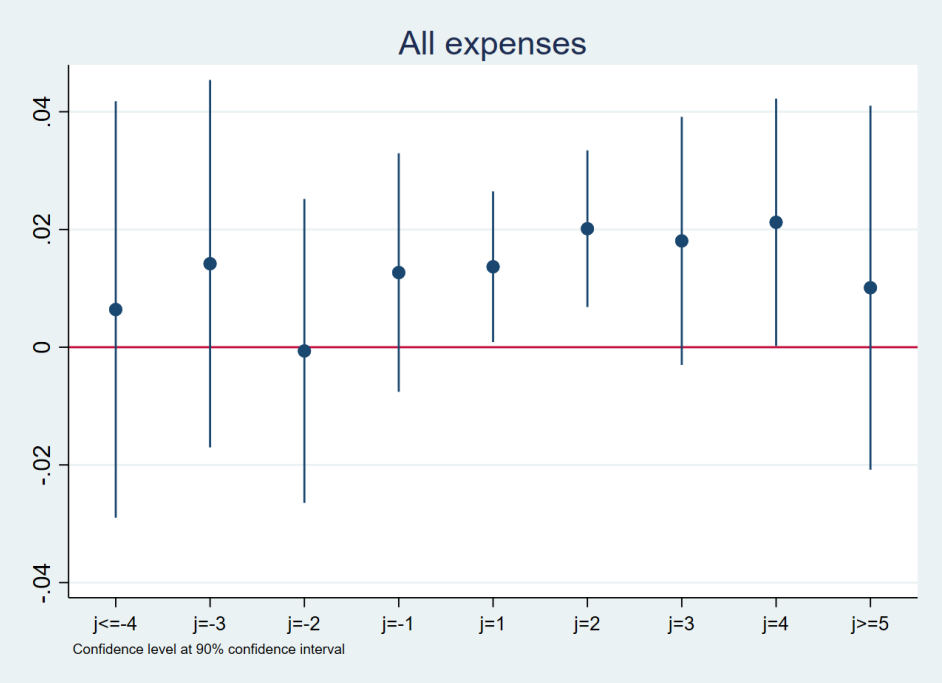
Table 3.6.5: Concordance with the local and national government pre and post 2010

| | (1) UKIP | (2) Green | (3) Other (ex. UKIP and Green) |
|---|---|---|---|
| Flooded | -0.586** | 0.201 | -0.0712 |
| | (0.240) | (0.226) | (0.729) |
| LocGov | -0.299 | 0.0830 | -1.492*** |
| | (0.252) | (0.179) | (0.474) |
| NatGov | -1.658*** | -0.102 | 1.447** |
| | (0.346) | (0.235) | (0.568) |
| Flooded × LocGov | 1.285*** | -1.085*** | -0.0751 |
| | (0.415) | (0.335) | (0.936) |
| Flooded × NatGov | 1.002** | 0.668** | -0.0330 |
| | (0.459) | (0.311) | (0.953) |
| Post | 5.644*** | 1.662*** | -3.402*** |
| | (0.324) | (0.238) | (0.585) |
| Flooded × Post | 2.356*** | 0.196 | -1.611 |
| | (0.820) | (0.443) | (1.086) |
| LocGov × Post | 1.736*** | 0.221 | 1.223** |
| | (0.365) | (0.223) | (0.576) |
| NatGov × Post | 0.187 | 0.192 | 0.444 |
| | (0.568) | (0.385) | (0.925) |
| Flooded × LocGov × Post | 2.665*** | 0.700 | 1.137 |
| | (0.986) | (0.523) | (1.286) |
| Flooded × NatGov × Post | 3.485*** | -1.449*** | 1.215 |
| | (0.998) | (0.486) | (1.312) |
| N | 8856 | 8856 | 8856 |
| Rsquared | 0.247 | 0.0655 | 0.0295 |
| Year FE | No | No | No |
| Ward FE | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election.
Council is a dummy for concordance of the ward's incumbent with the local council majority's party.
Government is a dummy for concordance of the ward's incumbent with the national government's party.
Post is a dummy for elections happening after 2010.

We first run an event-study type of analysis on spending at the local authority level before and after a flood. We use year flooded as the excluded category and add local authority and year fixed effects. We present in Figure 3.7.1 and 3.7.2 results for the services most likely to be impacted by a flood: all expenses and housing expenses. The coefficients for one to three years after a flood are positive and significant to the ten percent significance level. These results show an extended effect of floods on expenditures up to 3 years after a flood, but do not provide information on the effect of party affiliation.

Figure 3.7.1: All services spending after a flood

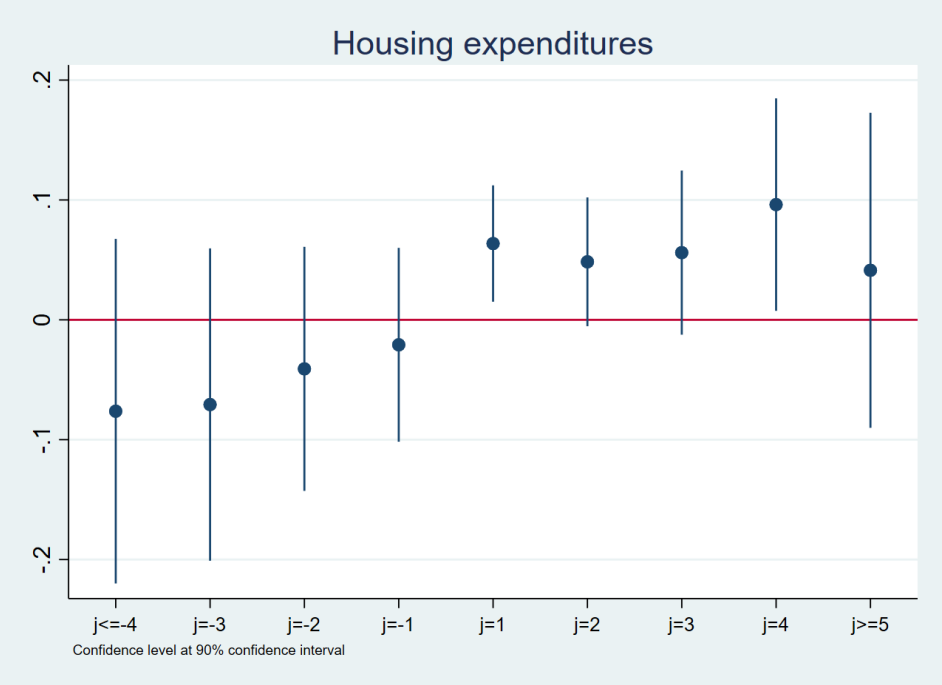



We turn to the impact of the alignment of the council with the national government and run the following specification:

$$ln(Expenditure_{ij}) = \omega_1 Flooded_{ij} + \omega_2 Alignment_{ij} + \omega_3 Flooded \times Alignment_{ij} + \psi_i + \rho_j + \upsilon_{ij}$$

Flooded is defined as flooded in any of the last 3 years (we show the results on spending

Figure 3.7.2: Housing spending after a flood




on housing by years before and after a flood in Figure 3.7.2 to motivate the long decay in flood impact on local expenditure). Alignment corresponds to alignment with the national government. $\psi_i$ and $\rho_j$ are local authority and year-fixed effects.

In Table 3.7.1, we find a slight increase in the overall expenditure by 4 percent, and a more considerable increase of 13 percent by looking at the subcategory of local housing (council housing is provided to households unable to afford private market housing or people that have been temporarily displaced).

Masiero and Santarossa (2020) found that in the context of earthquakes in Italy, the incumbent aligning with the national government successfully gains support after a disaster. They have linked those results to a greater ability of aligned local governments to provide relief. We thus look at this causal mechanism by interacting our flooded dummy with the alignment of the local authority with the national government in Table 3.7.2. While we found evidence that aligned local authorities increased their overall expenditures more, the sum of all coefficients is close to 0 – as aligned local authorities spent less in general. We do not find evidence of an increase linked to alignment for any interaction terms. Thus, we argue that

clientelism in expenditure after a flood is not a strong mechanism in England. Voters have thus little incentive to punish non-aligned incumbents and reward aligned ones. It is also concordant with the negative impact of being aligned with the local and national governments.

Table 3.7.1: Local expenditure after a flood

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | All services | Housing | Social care | Planning | Environmental |
| Flooded | 0.0419*** | 0.126*** | -0.0595 | 0.0324 | 0.0147 |
| | (0.0140) | (0.0417) | (0.0827) | (0.0272) | (0.0131) |
| N | 3443 | 3443 | 1899 | 3443 | 3443 |
| Rsquared | 0.946 | 0.508 | 0.556 | 0.785 | 0.932 |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| LA FE | Yes | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the local authority level
Councils are responsible for council housing, education, transport, planning, fire and safety, social care and waste among others. We present here results for areas most likely to be impacted by a flood.

Table 3.7.2: Local expenditure after a flood at the local council level

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | All services | Housing | Social care | Planning | Environmental |
| Flooded | 0.0225 | 0.142*** | -0.0686 | 0.0398 | 0.00253 |
| | (0.0139) | (0.0481) | (0.0776) | (0.0267) | (0.0150) |
| Government | -0.0310*** | 0.0726* | -0.00473 | 0.173*** | 0.0161 |
| | (0.0116) | (0.0412) | (0.0723) | (0.0269) | (0.0162) |
| Flooded × Government | 0.0494*** | -0.0410 | 0.0295 | -0.0158 | 0.0316* |
| | (0.0144) | (0.0627) | (0.0752) | (0.0303) | (0.0189) |
| N | 3443 | 3443 | 1899 | 3443 | 3443 |
| Rsquared | 0.946 | 0.509 | 0.556 | 0.799 | 0.933 |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| LA FE | Yes | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the local authority level
Councils are responsible for council housing, education, transport, planning, fire and safety, social care and waste among others. We present here results for areas most likely to be impacted by a flood.

## 3.8 Conclusion

In conclusion, we investigate the impact of floods on local elections. We find a negative impact on the incumbent party's share of votes. However, this effect is limited to floods within a year of the elections. Electors, however, strongly punish the incumbent if it is from the same party as the local or national government. These results differ from settings where the incumbents have benefited from natural disasters, primarily through their ability to mediate it with relief spending. On the contrary, in the UK, most of the decrease in the incumbent votes seemed to be captured by the protest party UKIP. There is also no strong evidence that local councils aligned with the government are likely to spend more after a flood.

This paper provides evidence that voters punish incumbents and turn to protest parties after natural disasters. Natural disasters can lead to a narrative of government failure. Reenforced by other socio-economic contexts such as austerity and Brexit, it can bolster protest parties and political changes. Local elections do not have the same ranging impacts as general elections or referendums – but they largely contributed to the rise of UKIP in the 2000s.

## Appendix

### 3.A.1 Data construction

Figure 3.A.1 shows the intersection of 2002 wards with wards from later periods. Figure 3.A.2 shows the intersection of wards and floods shapefile.

Figure 3.A.1: Wards intersection

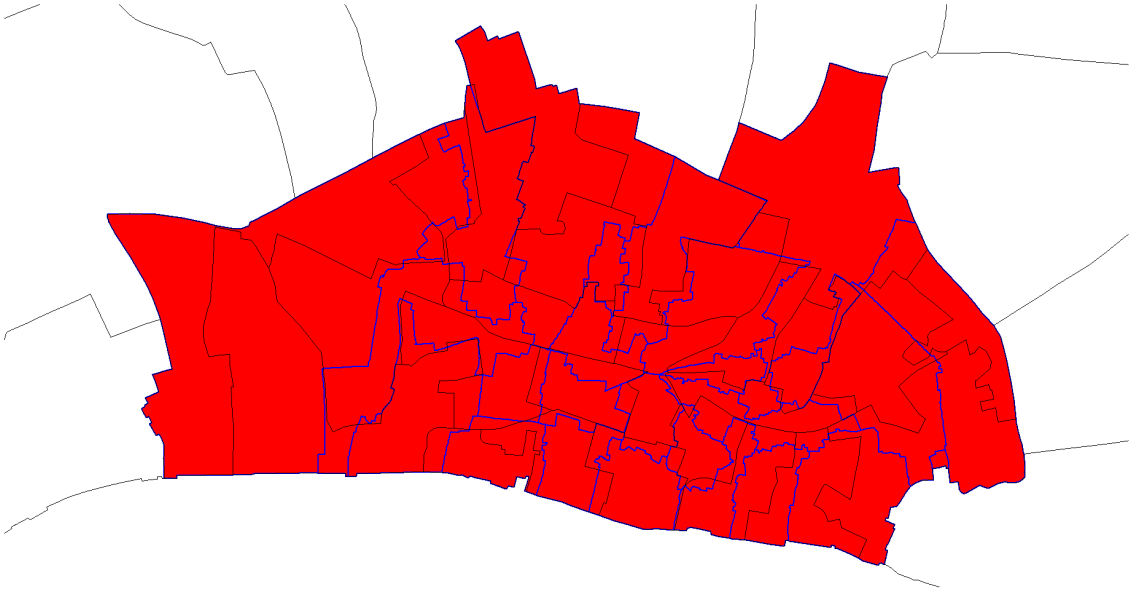
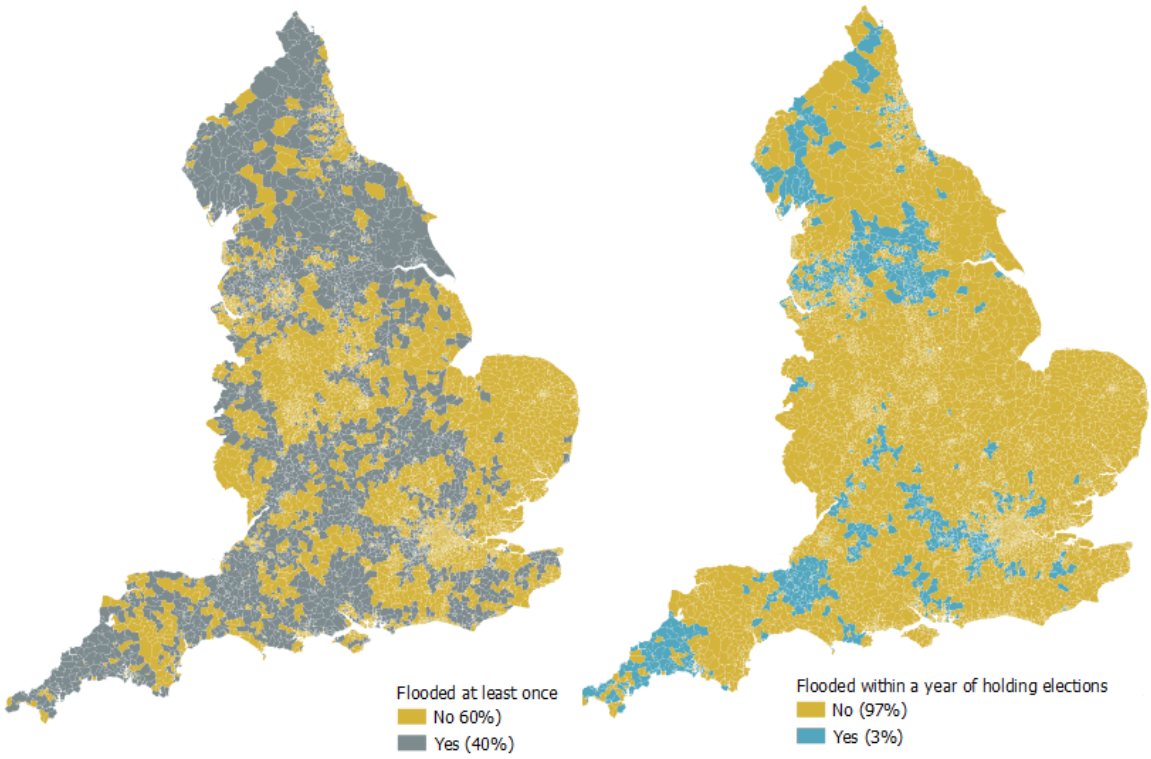


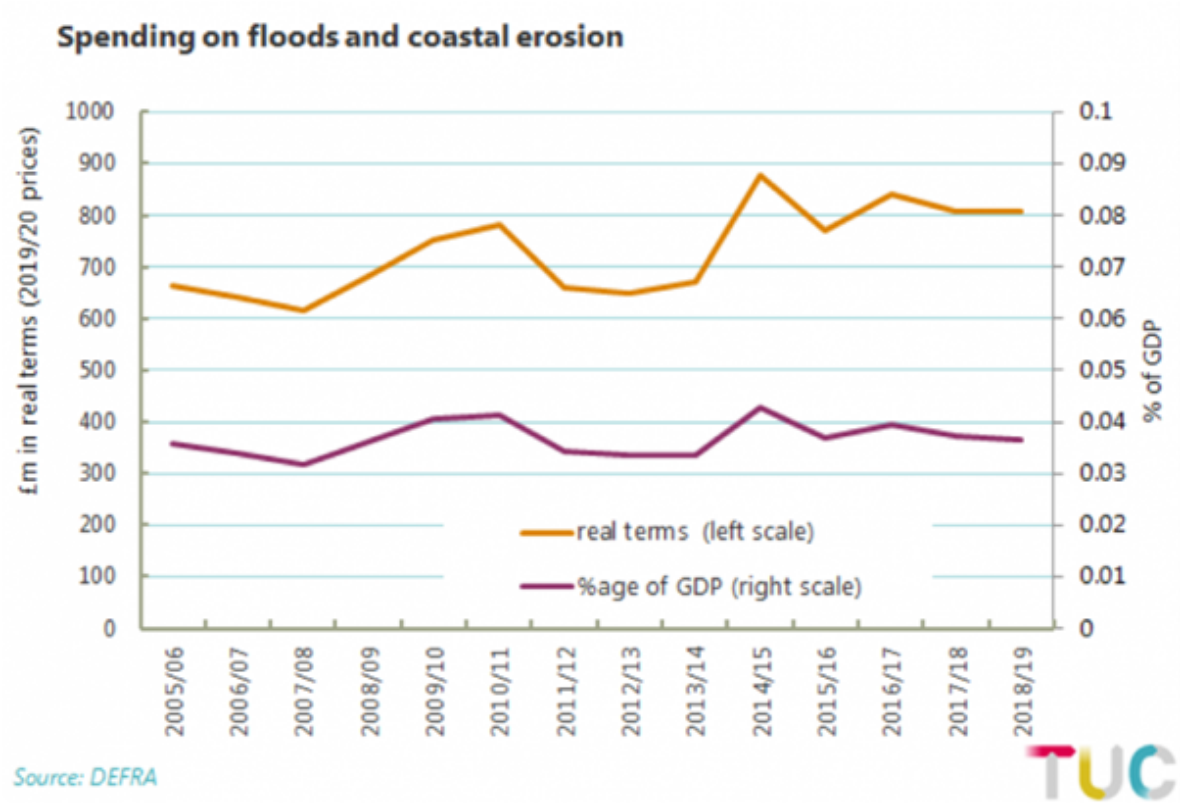
Figure 3.A.2: Flooded wards




### 3.A.2 Budget data

The following graphs show changes in spending and revenues at the local and national government levels.

Figure 3.A.3: Spending on floods and coastal erosion




### 3.A.3 Floods defences

The flood defences register maintained by the Environmental agency only shows a large increase in registered assets in 2012/2013 when it became mandatory to list assets in the database but has no new assets registered since (Figure 3.A.8).

Figure 3.A.4: Long-term trends in DEFRA spending

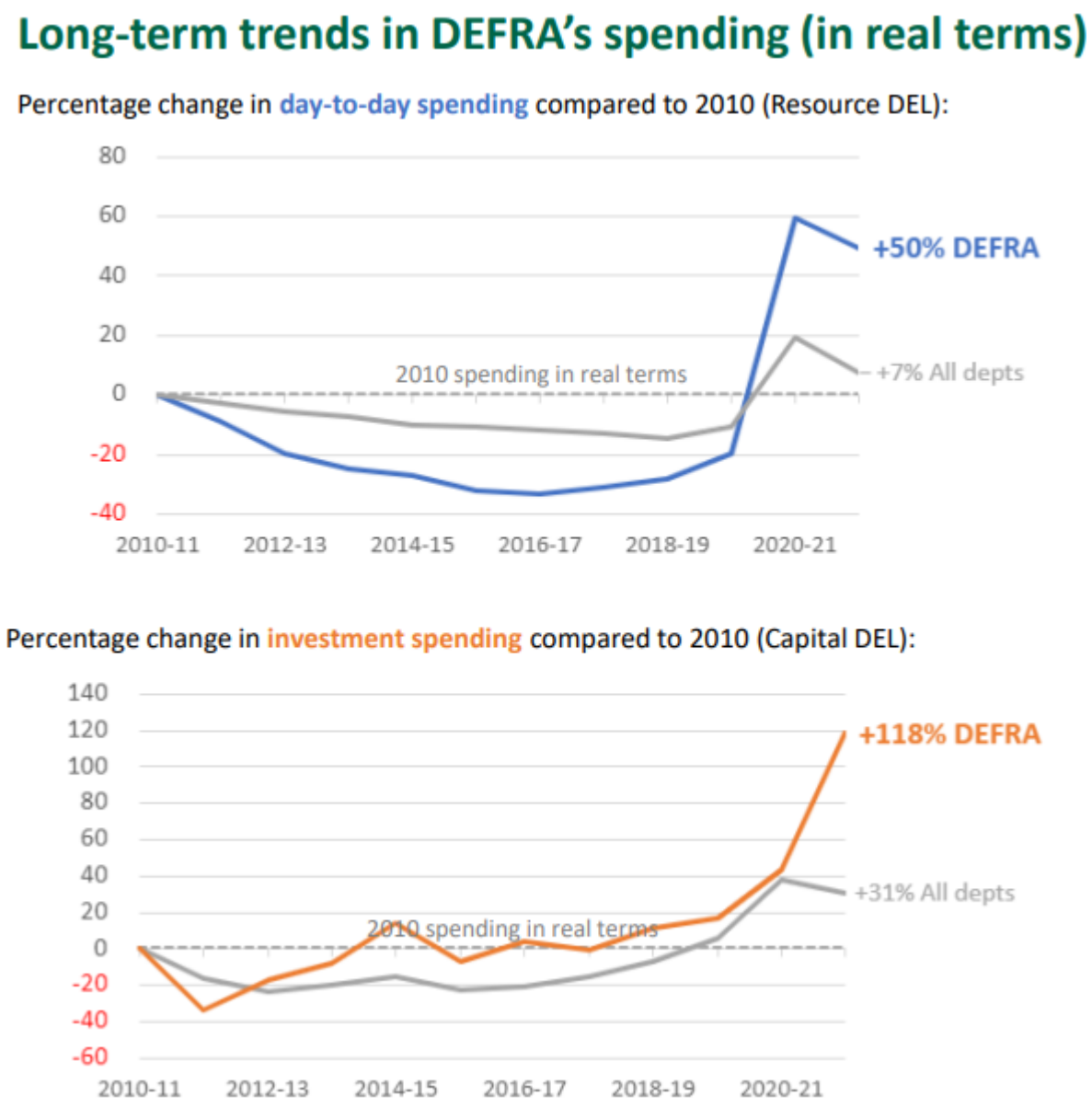


Figure 3.A.5: Local authority spending on flood defences in England

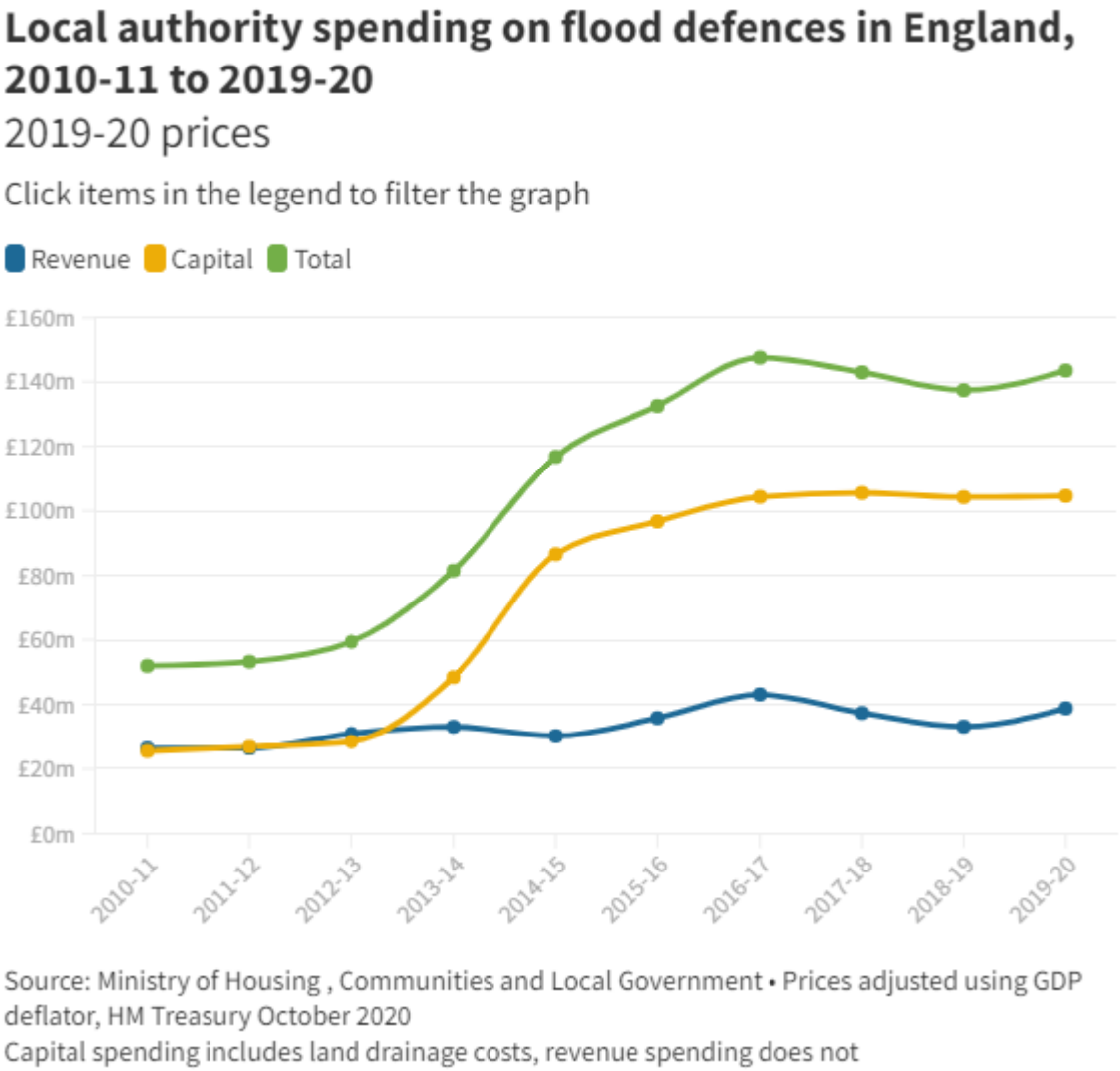


Figure 3.A.6: Local authority spending on coast protection in England

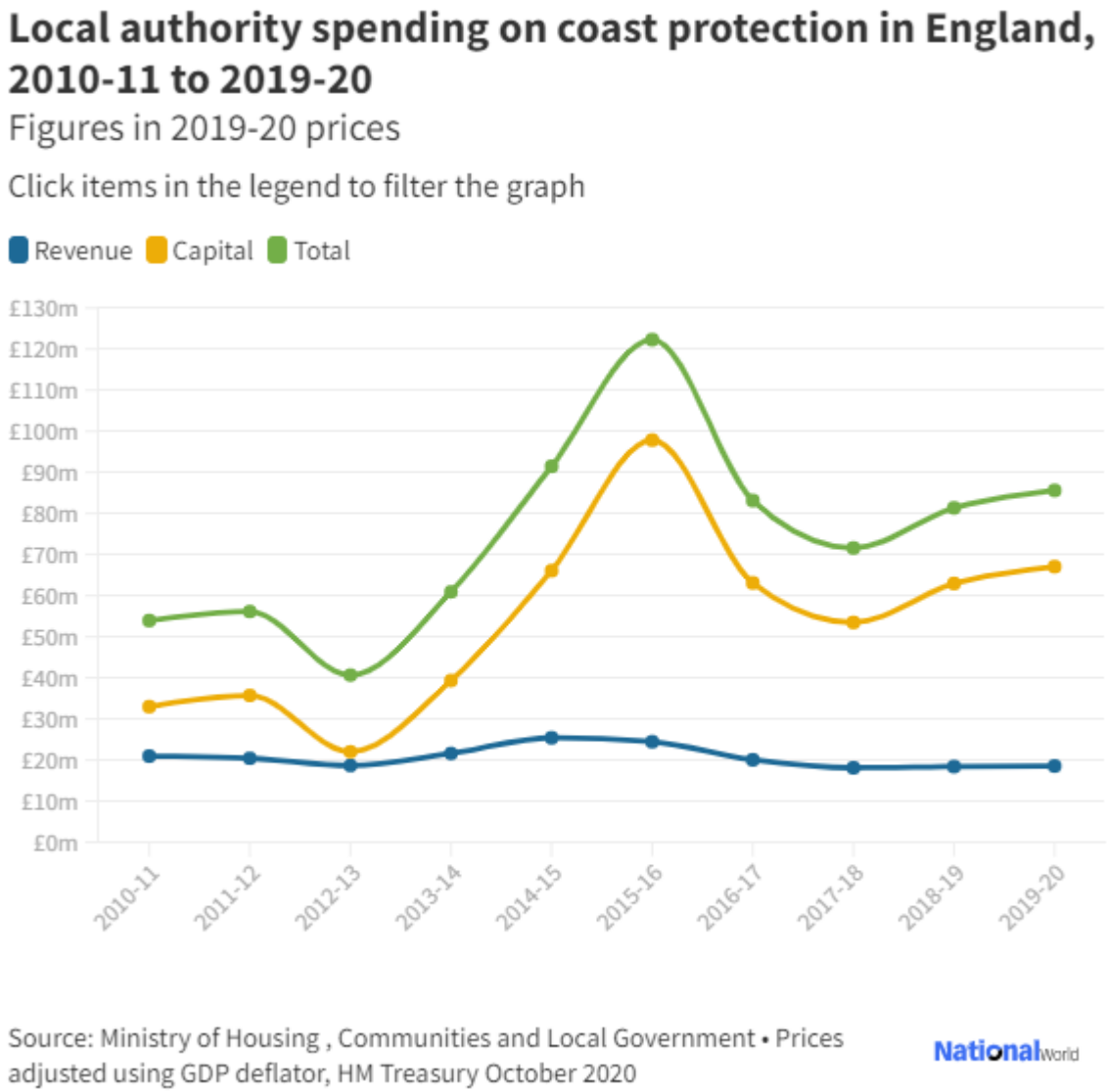
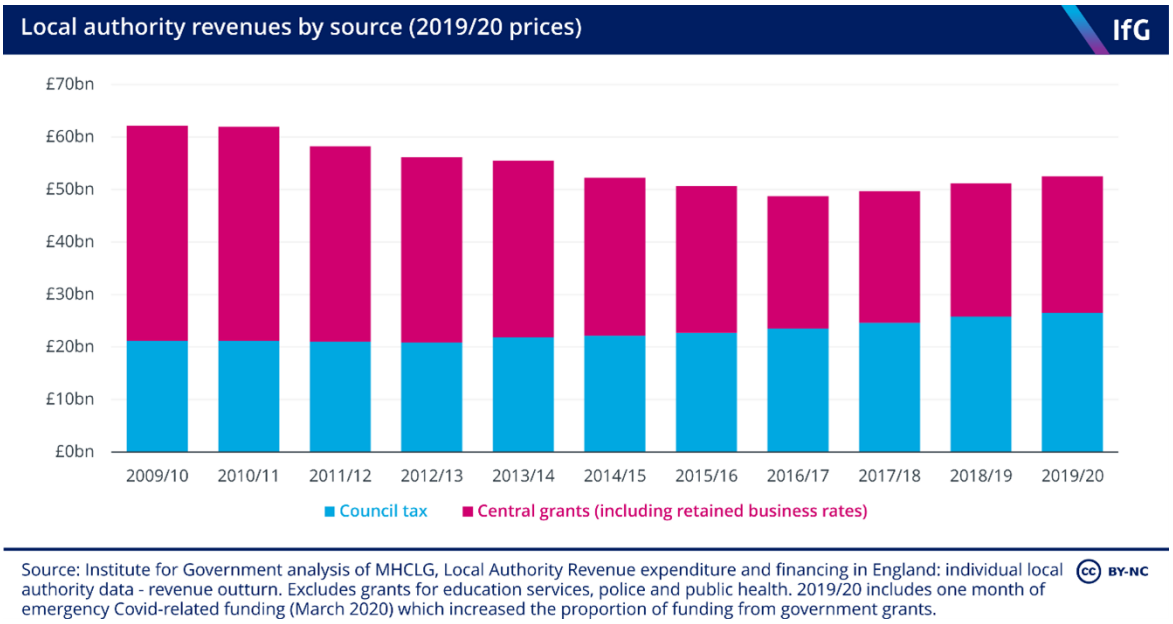


Figure 3.A.7: Local authority revenues by source




### 3.A.4 Additional tables

This section presents additional tables.

Figure 3.A.8: Floods defences registration

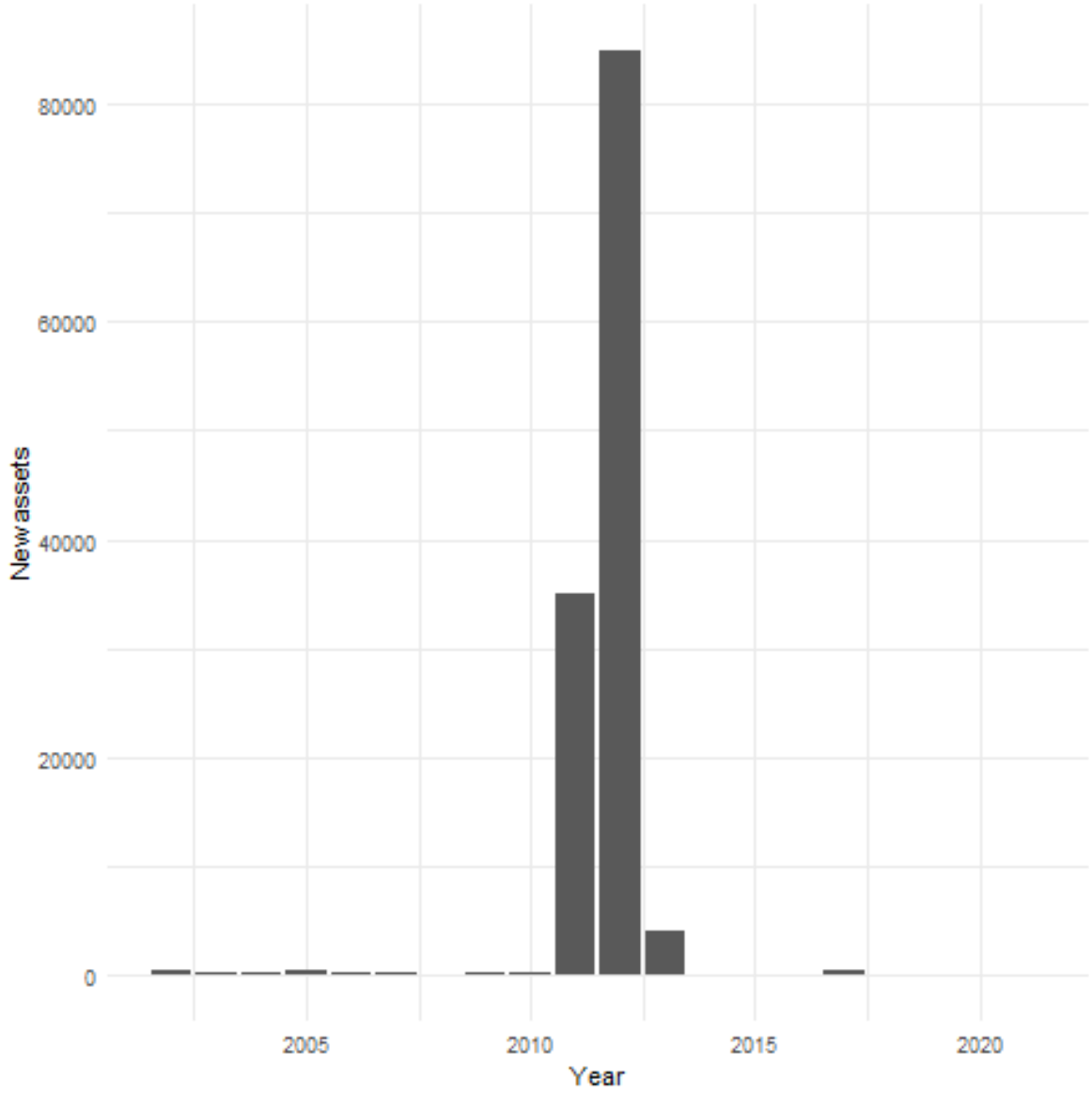


Table 3.A.1: Floods size and duration on the share of vote for the incumbent

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded | -0.949** | -0.133 | -0.0202 | 0.110 |
| | (0.428) | (0.488) | (0.490) | (0.489) |
| Flooded × Duration in days | -0.0276** | | -0.0181 | -0.0167 |
| | (0.0116) | | (0.0124) | (0.0128) |
| Flooded × Size of the flood (norm) | | -1.417*** | -1.322*** | -1.776*** |
| | | (0.361) | (0.372) | (0.431) |
| Flooded × Duration × Size | | | | 0.0384** |
| | | | | (0.0161) |
| N | 8856 | 8856 | 8856 | 8856 |
| Rsquared | 0.0720 | 0.0734 | 0.0736 | 0.0743 |
| Year FE | Yes | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election. Duration in days of the flood. Size of the flood is normalised using average flood size and standard deviation.

Table 3.A.2: Large and long-lasting events on the share of vote for the incumbent

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded | -0.492 | -0.737* | -0.387 | -0.190 |
| | (0.448) | (0.443) | (0.453) | (0.458) |
| Duration 75th | -3.473*** | | -2.996*** | -4.666*** |
| | (0.902) | | (1.070) | (1.327) |
| Size 75th pct | | -2.402** | -1.001 | -2.767* |
| | | (0.946) | (1.119) | (1.476) |
| Flooded × Duration × Size 75th pct | | | | 4.370* |
| | | | | (2.278) |
| N | 8856 | 8856 | 8856 | 8856 |
| Rsquared | 0.0734 | 0.0724 | 0.0735 | 0.0741 |
| Year FE | Yes | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election. The dummy variables duration and size 75th pct correspond to floods above the 75th percentile in terms of duration and area flooded.

Table 3.A.3: Concordance with the local government

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded | -1.234*** | | -1.163*** | -0.297 |
| | (0.413) | | (0.412) | (0.565) |
| LocalGov | | 3.753*** | 3.738*** | 3.976*** |
| | | (0.398) | (0.398) | (0.408) |
| Flooded × LocalGov | | | | -1.883** |
| | | | | (0.741) |
| N | 8856 | 8856 | 8856 | 8856 |
| Rsquared | 0.0715 | 0.0865 | 0.0875 | 0.0883 |
| Year FE | Yes | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election. LocalGov is a dummy for concordance of the ward's incumbent with the local council majority's party.

Table 3.A.4: Concordance with the national government

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded | -1.234*** | | -1.213*** | 0.0337 |
| | (0.413) | | (0.390) | (0.535) |
| NatGov | | -6.026*** | -6.025*** | -5.652*** |
| | | (0.297) | (0.298) | (0.311) |
| Flooded × NatGov | | | | -3.117*** |
| | | | | (0.728) |
| N | 8856 | 8856 | 8856 | 8856 |
| Rsquared | 0.0715 | 0.133 | 0.134 | 0.136 |
| Year FE | Yes | Yes | Yes | Yes |
| Ward FE | Yes | Yes | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election. NationalGov is a dummy for concordance of the ward's incumbent with the national government's party.

Table 3.A.5: LLFAs vs RMAs

| | (1) | (2) |
|---|---|---|
| Flooded | -0.840 | -0.861 |
| | (0.839) | (0.808) |
| NatGov | -5.540*** | |
| | (0.519) | |
| Flooded × NatGov | -2.004* | |
| | (1.128) | |
| LLFAs | 0 | 0 |
| | (.) | (.) |
| Flooded × LLFAs | 1.597 | 0.999 |
| | (1.021) | (1.069) |
| NatGov × LLFAs | -0.218 | |
| | (0.667) | |
| Flooded × NatGov × LLFAs | -2.022 | |
| | (1.455) | |
| LocGov | | 3.206*** |
| | | (0.598) |
| Flooded × LocGov | | -0.957 |
| | | (1.110) |
| LocGov × LLFAs | | 1.558* |
| | | (0.826) |
| Flooded × LocGov × LLFAs | | -1.681 |
| | | (1.495) |
| N | 8856 | 8856 |
| Rsquared | 0.136 | 0.0890 |
| Year FE | Yes | Yes |
| Ward FE | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election.
LLFAs is a dummy for Lead Local Flood Authorities established in 2010.

Table 3.A.6: One-tier vs Two-tier authorities

| | (1) | (2) |
|---|---|---|
| Flooded | -0.834 | -0.982 |
| | (0.818) | (0.778) |
| NatGov | -5.138*** | |
| | (0.483) | |
| Flooded × NatGov | -2.818** | |
| | (1.096) | |
| Unitary | 0 | 0 |
| | (.) | (.) |
| Flooded × Unitary | 1.686* | 1.333 |
| | (1.006) | (1.061) |
| NatGov × Unitary | -1.072* | |
| | (0.636) | |
| Flooded × NatGov × Unitary | -0.362 | |
| | (1.440) | |
| Flooded × LocGov | | -1.656 |
| | | (1.076) |
| LocGov × Unitary | | 0.341 |
| | | (0.820) |
| Flooded × LocGov × Unitary | | -0.348 |
| | | (1.488) |
| N | 8856 | 8856 |
| Rsquared | 0.137 | 0.0886 |
| Year FE | Yes | Yes |
| Ward FE | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election.
Unitary is a dummy for One-Tier local authorities.

Table 3.A.7: Number of times flooded

| | (1) | (2) |
|---|---|---|
| Flooded | 1.592 | 0.511 |
| | (1.080) | (1.130) |
| NatGov | -6.863*** | |
| | (0.604) | |
| Flooded × NatGov | -3.450** | |
| | (1.533) | |
| Number of times flooded | 0 | 0 |
| | (.) | (.) |
| Flooded × Number of times flooded | -0.687 | -0.347 |
| | (0.449) | (0.462) |
| NatGov × Number of times flooded | 0.634** | |
| | (0.293) | |
| Flooded × NatGov × Number of times flooded | 0.0819 | |
| | (0.665) | |
| LocGov | | 3.033*** |
| | | (0.859) |
| Flooded × LocGov | | 0.0814 |
| | | (1.564) |
| LocGov × Number of times flooded | | 0.483 |
| | | (0.398) |
| Flooded × LocGov × Number of times flooded | | -0.968 |
| | | (0.657) |
| N | 8856 | 8856 |
| Rsquared | 0.137 | 0.0893 |
| Year FE | Yes | Yes |
| Ward FE | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election.
Number of times flooded is the number of times a ward got flooded in the last 30 years.

Table 3.A.8: Close elections

| | (1) | (2) |
|---|---|---|
| Flooded | -0.0178 | -0.0574 |
| | (0.627) | (0.720) |
| Flooded × NatGov | -2.689*** | |
| | (0.938) | |
| Close elections | -8.409*** | -7.001*** |
| | (0.367) | (0.393) |
| Flooded × Close elections | -0.521 | -0.673 |
| | (0.888) | (0.916) |
| NatGov × Close elections | 2.499*** | |
| | (0.486) | |
| Flooded × NatGov × Close elections | 1.068 | |
| | (1.304) | |
| LocGov | | 4.569*** |
| | | (0.504) |
| Flooded × LocGov | | -1.494 |
| | | (0.919) |
| LocGov × Close elections | | -2.723*** |
| | | (0.547) |
| Flooded × LocGov × Close elections | | 0.161 |
| | | (1.302) |
| N | 8856 | 8856 |
| Rsquared | 0.207 | 0.177 |
| Year FE | Yes | Yes |
| Ward FE | Yes | Yes |

* $p < .1$, ** $p < .05$, *** $p < .01$
Standard errors clustered at the ward level
The sample is all wards flooded within a year of an election.
Flooded is a dummy for flooded the year before a local election.
Close election is a dummy for average difference between first and second party of less than 10 points.

# 4 Nowhere else to go? Urbanisation and Flood Risks: The role of land scarcity

## 4.1 Introduction

The world urban population share went from 40% of the world population in 1985 to 54% in 2016 and a projected 68% by 2050 (Ritchie and Roser 2020). A large share of the population is moving into cities, expanding existing settlements and creating new ones. This urban growth is occurring in areas that have been neglected before, including land exposed to flood risks such as riverbeds, flood plains, or wetlands. This study explores the trends in urban expansion and flood exposure in China over a 30-year window from 1985-2015.

In the last thirty years, settlement growth in high flood risk zones has consistently outpaced growth in no- and low-risk zones (Rentschler et al. 2022). In addition to building in flood prone areas, cities are also re-building on hazard-prone land after a climate disaster (Lin, McDermott, and Michaels 2021). The magnitude of this problem is vast. The amount of newly settled land in high flood risk areas over our period of analysis totals 36,500 km2, or three times the size of the New York City Metropolitan area (Rentschler et al. 2022). People living in high flood risk zones around the world totalled 1.47 billion in 2020 (Stéphane Hallegatte et al. 2020a). This exposure resulted in 650,000 fatalities and another 650 million displaced over the 1982 to 2014 period (G. Brakenridge 2016). Under recent projections, this trend will only accentuate, with an at least five-fold increase in the population that experience coastal flooding annually over the next century (Adger, Arnell, and Tompkins 2005).

The causes of urban expansion in high flood risk areas are varied: settlements are physically constrained, face tight urban planning restrictions or a fast-growing population that overwhelms urban planning. In this paper, we focus on the impact of land scarcity due to geographic obstacles: the share of unviable land (e.g. land that is very costly to develop on) surrounding a city due to geographic obstacles (steep land, rivers, and oceans) or high flood risk. We do not include land-use policy induced land scarcity (e.g. zoning policies).

Our paper studies settlement expansion in high flood-risk land in China. Home to over half

of the high-risk settlement growth over the 1985 to 2015 period (Rentschler et al. 2022), China is a particularly important context for studying settlement growth in high flood risk. China's rapidly expanding population is associated with urban sprawl, booming urban development and strict land use regulations, which makes risky urban growth an important policy debate in China.

Are settlements increasingly expanding into high flood risk land due to the lack of viable and low flood risk land surrounding existing settlements? Using satellite data and a spatial instrumental variable, we empirically estimate whether land scarcity is a driver of city expansion into high flood risk land.

At face value, geographic obstacles appear to be exogenous constraints. However, the main contribution of this paper is to consider the empirical challenge to do with the endogeneity of land scarcity: geographic obstacles are not random and result from the selection of urban planners to settle and expand in their proximity to benefit from the consumption and amenity value of that land. Indeed, settlers founded cities near water bodies because they valued their inherent consumption, amenity and economic value: such as the proximity to trade routes and fertile land[12]. Settlements continue to expand along water bodies and mountains for the inherent amenity value of the land and due to path dependency of settlement growth. An important corollary to this is that settlements near water bodies might grow faster than safer settlements situated in more remote land.

Using US data, Lin, McDermott, and Michaels (2021) show that cities first developed on safe land near the coast and then increasingly expanded towards riskier land. This was also demonstrated in Vietnam, where the safest and most productive locations were occupied first, and left new developments forced to use sub-optimal land (Rentschler and Salhab 2020). The relationship between urban expansion in high flood risk areas and land scarcity is thus an equilibrium outcome of diverging forces: preferences for safe areas, path dependence of city expansion, productivity and amenity value of risky land (floodable land (e.g. water bodies) and steep land (e.g. mountains)).

[12]and beneficial terrain (ruggedness or bedrock depth)

To test whether land scarcity is a driver of urban growth in high flood risk land, we need to isolate the role of land scarcity from the endogenous forces described above. We propose a novel instrument for changes in land scarcity that combines geography with a mechanical model for city expansion and removes the endogeneity of voluntary city expansion towards water bodies or mountains.

The underlying idea is that, as cities expand in space and over time, they face geographic constraints— steep terrain or bodies of water — that limit the amount of viable land they can expand into. The initial city location in space is not random - for example, being close to mountains affects a city's past economic specialization, resilience and growth. However, the relative position in space of such geographic constraints as a city expands allows for available land in cities' vicinity to vary randomly. We build an instrument to capture this variation. The identification relies on changes in available land in the vicinity of cities as a result of a city randomly hitting geographic obstacles.

While land scarcity induced by geographic obstacles appears to play a large role when looking at cities statically, we provide causal evidence that, once we remove the impact of original settlement choice, the increasing 'bite' geographic barriers such as mountains or water bodies are not a strong driver of urbanisation in high flood risk areas.

In other words, cities do not expand into high flood risk land because they have nowhere else to go. We show that expansion of cities in high flood risk areas is largely due to path dependency – it is mostly happening in cities that already face high flood risk. As global warming is predicted to exacerbate the frequency and magnitude of flooding, more evidence is needed on the dynamics of urban expansion into high flood risk land. Our results are important as they point to the path dependence of cities towards hazard-prone land, even when safer land is available.

We first review the literature in Section 4.2 and present the data in Section 4.3. Next, we present our estimation strategy and instrument for changes in land scarcity in Section 4.4 and

Section 4.5. We then discuss the results when we separate out the role of land scarcity per se from the amenity value of geographic barriers in Section 4.6 and conclude in Section 4.7.

## 4.2 Literature review

This paper contributes to three strands of literature: i) where and why are populations settling in risky areas, specifically in flood-prone land, ii) drivers of city size and urban sprawl, iii) urban structures in response to shocks.

Faced with planning restrictions and lack of space, low-income households have moved to neighbourhoods plagued with environmental risks. Case-study evidence has illustrated these mechanisms. In Mumbai, India, low-income households have moved into the city from rural peripheral regions in search of economic opportunities. The high density of existing settlements and land price differentials have forced new migrants into areas that have been avoided in the past, such as high-risk land in the proximity of riverbeds that lack planning and public infrastructure such as drainage or flood defence systems. Similarly, informal neighbourhoods have lacked access to public services such as sanitation, drainage or flood defence systems, such as the steep hills of Rio de Janeiro or the São Paulo favelas. They are costly to formalise or upgrade, leading to lower population density and land values (Harari 2020). As a result, settlement growth in high flood risk zones has consistently outpaced growth in no- and low-risk zones over the last 30 years (Rentschler et al. 2022). An extensive literature review can be found in (Stéphane Hallegatte et al. 2020b) and they conclude that land and housing markets often push poorer people to develop settlements in riskier areas, especially where land is scarce.

Evidence for this can be found around the world: Daniel, Florax, and Rietveld (2009) conduct a meta-analysis of the literature and find that a 1 percentage point increase in the annual probability of flooding is associated with a 0.6 percent decrease in house prices. In Accra, Ghana, Erman et al. (2018) find that flood-affected dwellings are valued at 30 percent less on average than unaffected ones. In Dar es Salaam, Tanzania, households affected by floods lost on average 23 percent of annual income and self-evaluated their dwellings to be worth about 36 percent less than non-flood-prone dwellings (Erman et al. 2019). Reduced housing prices, in combination with informal housing markets, make it possible for poor people to

access housing opportunities that could be out of reach in the absence of risk Durand-Lasserve, Durand-Lasserve, and Selod (2015). Additionally, in such places, it may not only be the prices that push poor people into risky places but simply the availability of land with appropriate access to jobs and services. From Pune, Dhaka, Caracas, Rio de Janeiro to Mumbai, informal settlements are often located in hazard-prone locations such as on hillsides, close to riverbanks, or near open drains and sewers Lall, Lundberg, and Shalizi (2008). In Chapter 5, I show that in Sub-Saharan Africa, households living in areas that are repeatedly flooded are systematically poorer and less healthy than households in safer areas – however, the difference in wealth and health between rural and urban areas is such that households living in dangerous urban areas still have better outcomes than in safe rural areas. Our paper contributes to a large literature on urban sprawl (Glaeser, Gyourko, and Saks 2006). Our work follows Burchfield et al. (2006) in using remotely sensed data to track the urban sprawl of cities. Our instrumental variable approach is built on identifying assumptions developed in a paper by Harari (2020), who studied how geographic barriers around Indian cities impact city shape and, in turn, economic outcomes, using night-time data imagery.

Our paper also contributes to a separate strand of papers in urban economics, focused on path dependence in the spatial distribution of economic activity and urban settlement and the relationship between geography and growth (Desmet and Henderson 2015). As urban areas are predominately in floodable areas, the increase in urbanisation would increase the share of population at risk, even without cities expanding into higher risk land. Combining satellite-based nightlight imagery and flood hazard maps, Desmet et al. (2021) and later Rentschler and Salhab (2020) systematically assessed risks in high-growth areas. Their analysis confirms that areas with high urban and economic growth face significantly higher flood risk than low-growth areas. About 27 percent of areas with low urban and economic growth are estimated to be exposed to flooding with a 100-year return period, compared to some 50 percent of high-growth areas. Also combining satellite-based nightlight imagery and large flood events which affected 1,868 cities in 40 countries around the globe, Kocornik-Mina et al. (2018) document a lack of adaptation or movement of economic activity away from the most flood-prone locations within cities. We add to this literature by testing the amount of path dependence at the city level by initial exposure to floodable land. Recent papers have studied

the costly path dependence of cities in relation to sea level rise. Balboni (2021) studies the exposure of Vietnam roads to sea level rise and finds that infrastructure investments that ignore future sea-level rise risks might lead to inefficient persistence in coastal cities. Desmet et al. (2021) use a spatially disaggregated, dynamic model of the world economy to quantify the role of migration and local agglomeration in the projection of sea-level rise cost. Lin, McDermott, and Michaels (2021) explore the internal structure of coastal cities and their adjustment to climate change. They characterise "soft" barriers, such as flood-prone areas, as locations that are not used for housing development in most circumstances but are nevertheless built on as cities expand. They document how new construction in the U.S. in recent decades avoided flood-prone areas in sparse locations but did take place on (the 'least-bad') flood-prone areas in dense locations. They develop a monocentric coastal city model where flood-prone settlement results from the trade-off between the amenity of coastal proximity and the disamenity of flood risk.

Our paper is the first to test the causal role of increased land scarcity and path dependence on urban growth in high flood risk land.

## 4.3 Data

### 4.3.1 Settlement growth

We use the World Settlement Footprint-Evolution developed by the German Aerospace Center - DLR (Marconcini et al. 2021) to study settlement expansion. This new dataset combines the best of Landsat-8 multispectral satellite images and Corpernicus Sentinel-1 satellite data to offer a 30-metre resolution binary mask outlining the 2015 global settlement extent. From the 2015 snapshot, the DLR created a yearly panel of settlement extent from 1985 to 2015 using backward iterative techniques. From this satellite imagery, we create a panel dataset of cities around China from 1985 to 2015.

One limitation of previous datasets such as Landsat is the poor quality and patchy coverage of satellite images for the 1980s and 1990s in specific regions such as Sub-Saharan Africa[13].

[13]In these cases, the frequency of satellite images was lower, thus reducing the likelihood of high-quality cloud-free images being available in a certain year.

The dataset has been extensively ground-truth tested with 900.000 validation samples to correct this limitation.[14]

The high resolution enables tracking not only of large cities, but also of small settlements[15]. We define cities as any continuous patch of settlement extent in 1985 that is larger than 1 km square. We do not differentiate between urban and rural settings as in some other datasets (eg. GHSL)[16].

We then overlap the continuous patch of settlement extent of more than 1 km square in 2000 and 2015. In our analysis, we only keep settlement footprints that intersect 1985 settlements to capture the impact of scarcity using a long-difference between two 15-year periods (1985-2000 and 2000-2015). A limitation of our estimation strategy is that we cannot study new cities that appear between 1985 and 2015.

In the context of China, it is common for cities to merge into one continuous urban area over our period[17]. In order to keep a stable unit of analysis, we consider the non-contiguous patches that will merge into one by 2015 as one unit of analysis from 1985 onwards. To calculate the area growth of that unit of analysis, we sum up the areas of the separate patches. We proceed similarly for all the measures we construct. We include an illustration in Figure 4.A.1 in the Appendix.

City growth can occur in three ways: edge expansion, infilling, or leapfrogging. In this paper, we focus on the edge expansion of the city. We define edge expansion as any new pixels of settlement contiguous to the city shape in 1985. We focus on the extensive margin as our data does not allow us to measure the intensive margin: once a grid square is built, we do

[14]As part of overcoming any limitations of settlement over-or under-estimation (due to limited satellite scene collection in specific countries before 2000), a cutting-edge regression model has been applied to model the settlement extent using data availability specific to the region. The DLR also provides a novel index to measure the quality of the output at the 30m resolution, which allows us to conduct robustness checks removing low-quality data.

[15]Note that the settlement labelling protocol is defined on a taxonomy of buildings, building lots and roads/paved-surfaces.

[16]The Global Human Settlement Layer (GHSL) dataset, for example, applies a cut-off of 300 inhabitants per km square and a minimum of 5,000 inhabitants to distinguish between urban and rural areas.

[17]In our data, this corresponds to a continuous patch of settlement in 2015 which overlaps several non-contiguous patches in 1985 and 2000.

not observe the increase in density. As we observe cities in our sample from 1985 to 2015 and discretise the period into two 15-year periods, we can calculate expansion and infilling for 2000 and 2015[18], and leapfrogging for 1985 and 2000[19]. In 2000, we find that 65% of new settlements correspond to edge expansion, against 30% for leapfrogging and 5% for infilling. Focusing on new settlements in high flood risk areas, we find similar proportions: 67%, 30% and 3% for each type respectively. To study the impact of land scarcity, we need to define what is unviable land. Following Harari (2020) and Saiz (2010), we define unviable land as any water body, ocean or steep terrain (above 15%). The steepness is calculated from DEM MERIT elevation model. Water bodies are extracted from the OSM water layer from IIS U-Tokyo.

### 4.3.2 Floods

We look at three types of floods: fluvial, pluvial and coastal. Fluvial flooding occurs when water bodies overflow onto adjacent land due to precipitation or snow melt. Pluvial flooding occurs when the absorptive capacity of the soil is exceeded. It is common on impervious surfaces in urban areas or after droughts. The fluvial and pluvial flood maps are from the 2019 Fathom-Global 2.0 dataset (Sampson et al. 2015). Fathom is a global flood model that uses terrain and hydrological data to predict flood risk probability and flood risk depth at 90-meter resolution for the entire world. The Fathom-Global 2.0 uses the newest DEM MERIT elevation model that corrects for multiple errors, including tree and building height adjustments. The Fathom flood models have been shown to have good predictive performance. The most comprehensive performance metric used in the analysis of flood models is the critical success index (CSI), which measures model fit by measuring the share of total forecasts which are correct. Bernhofen et al. (2018) find that the Fathom-Global 2.0 scores highly on this measure, with a score over 0.7 CSI in case studies. The flood hazard maps also perform better than the climate-forced models for average flood return rate (Bernhofen et al. 2018). In our sample, 67% and 70% of settled pixels in high flood risk areas correspond to fluvial risks in 2000 and 2015 against 25% and 33% for pluvial flood risks and only 4% and 3% for coastal

[18]We observe the stock of pixels in 1985, so we cannot distinguish pixels that were added as the city expanded or plots that were surrounded by built-up before being infilled.

[19]The difference between leapfrogging and edge-expansion depends on the number of years included in each period. We use 1985-2000 and 2001-2015 as it allows for a period long enough to distinguish large changes in settlement while still allowing for the long-difference analysis. We cannot know which surrounding pixels will belong to the continuous city "core" in 2030, so we cannot calculate leapfrogging for this period.

flood risks[20]. The model simulates flood events with return periods of 5, 20, 50, 100, 250, and 500 years.

Coastal flooding is caused by storm surges and high tides in coastal areas. We use 3-degree resolution (~90 m at the equator) Joint Research Centre (JRC, 2014) coastal flood risk maps developed by the European Commission, which was previously used in a global-scale analysis (Koks et al. 2019). These coastal inundation maps are simulated using the LISFLOOD-FP hydrological model Vousdoukas et al. (2016). Coastal flood simulations are forced by extreme sea-level rise derived from reanalysis of waves and storm surges (Muis et al. 2016), and further combined with tidal information (Vousdoukas et al. 2018). A high-accuracy spaceborne digital elevation model, in which absolute bias, stripe noise, speckle noise, and tree height bias are corrected, is used as an input to the model (Yamazaki et al. 2017).

These flood hazard maps provide both a measure of flood severity – measured by the potential inundation depth of a given flood - and of flood probability, also named return rate or periods. A return period describes how much time will pass before the next flood of the same intensity occurs again. Our main analysis uses the 1 in 100-year return rate, where the flood is expected to occur once every 100 years (i.e., it has a probability of 1% of occurring in any given year). On average, this return rate also means there is a 10% chance of a flood occurring in a decade, or 50% in a lifetime (68 years). These are significant probabilities that lie within reasonable planning horizons of governments, and that are widely used among development banks and flood insurance companies.

In terms of flood severity, the flood hazard maps include potential inundation depth on a continuous scale. We aggregate the values into five categories to reflect the risk to human life. The 'no risk' category refers to areas unaffected during a 1-in-100-year flood. Up to 0.15 meters inundation depth, no significant risk to life is expected ('low risk'). Up to 0.5 meters, some risk to life must be expected, especially for vulnerable groups such as children and the disabled ('moderate risk'). Up to 1.5 meters, a significant share of the affected population could face risk to life, especially if flood waters have a current ('high risk'). Above 1.5 meters,

[20]Total sum adds to more than 100% as a pixel can be subject to several flood risks.

most affected people could face substantial risk to life without rescue measures ('very high risk'). We consider flood risks to be significant when inundation depth is higher than 0.5 meters (high risk to 'very high' categories). For our main results, we select the very high flood risk category (henceforth HFR), where inundation depth is higher than 1.5 meters. We conduct heterogeneity analysis for lower inundation depths.

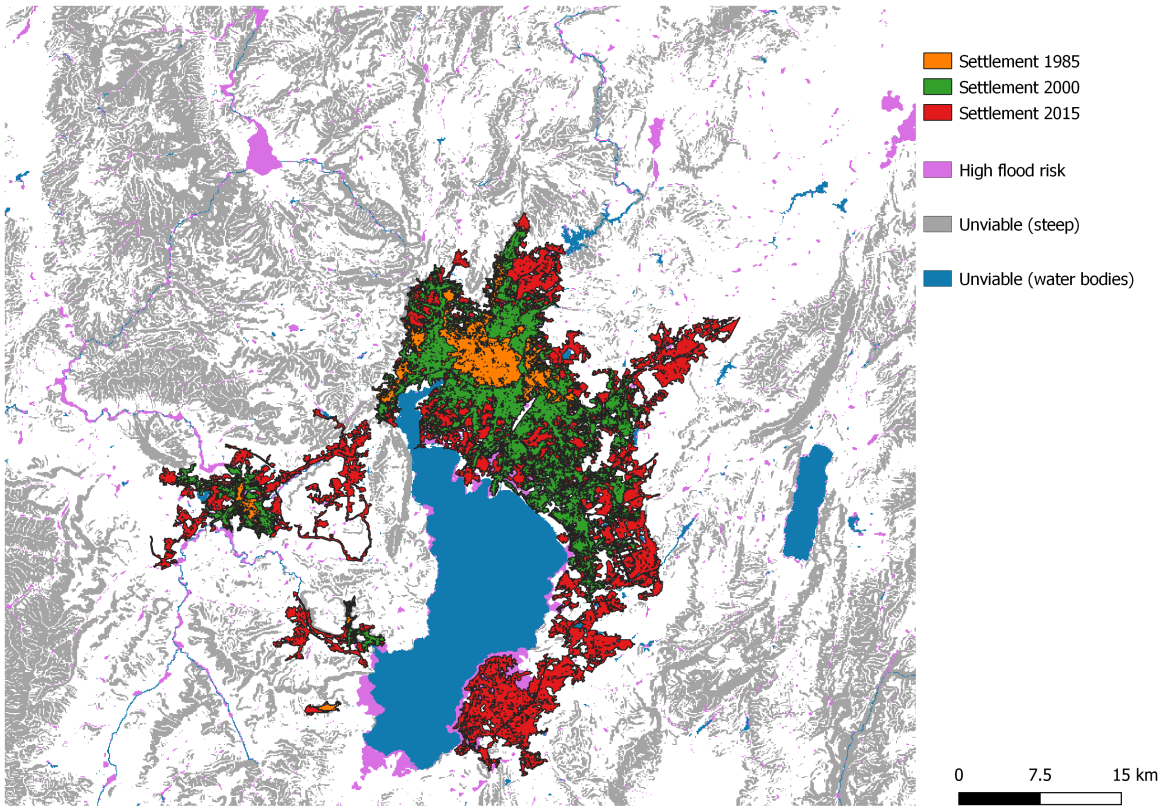
As with all existing global flood maps, the effects of artificial flood protection structures like dikes are not incorporated. This data limitation is pervasive in this literature as no complete global inventory of flood defences exists. Ongoing initiatives, such as the FLOPROS database, could eventually fulfil this need but are currently still falling short of comprehensive coverage (Scussolini et al. 2016). While 'defended' hazard maps are available, they rely on models that proxy the likelihood of flood protections given observable characteristics such as economic prosperity or population density, thereby introducing biases for our analysis of flood risk and urbanisation. Looking at un-defended flood models will lead us to overestimate the share of land in floodable areas where flood defence infrastructures were developed. This might suggest that we will see different effects for richer and poorer countries. For our empirical investigation, we overlay the settlement extent data, flood maps and elevation data to label each settled pixel as high flood risk or unviable. We also discretise our analysis to two 15-year time periods 1985-2000, and 2000-2015. The resulting data output is Figure 4.3.1.

## 4.4 Empirical investigation: the role of land scarcity

To frame the empirical investigation of the impact of land scarcity on settlement growth in high flood risk land, we extend the classic monocentric city model (Appendix Section 4.A.2). We show that the city will move further away from the monocentric model as a function of the size and location of unviable land. We also show that population will settle in high flood risk land if they value the amenity value of the land over the dis-amenity value of the flood, given the commuting benefits associated with the land.

A major driver of urbanisation in high flood risk areas is the scarcity of viable and safe land around a city (Stephane Hallegatte et al. 2017). Are cities expanding into high flood risk

Figure 4.3.1: Map of the 30m spatial resolution WSF evolution layer, overlayed with the flood hazard maps, for a city in China



because they have nowhere else to go? In this section, we formally test this hypothesis.

Empirically, we want to estimate whether increasing land scarcity is indeed driving cities to develop more on high flood risk land compared to unconstrained cities. Our dependent variable is urban growth in high flood risk areas. We measure this as the share of new built-up settlement extent (SE) of a city $i$ built in high flood risk (HFR) areas between periods $t-1$ and $t$ ($Y_{t,i}$). High flood risk (HFR) areas are defined as zones with an inundation depth above 1.5 meters using 1 in 100 years flood risk return rate (such depth is a danger to human life). At face value, this investigation is straightforward – we want to measure "naively" how constrained a city is by looking at the amount of unviable land around the city. Relating to Figure 4.3.1, our outcome is the change in the share of HFR pixels between the dark green settlements of 2000 and the red settlements of 2015.

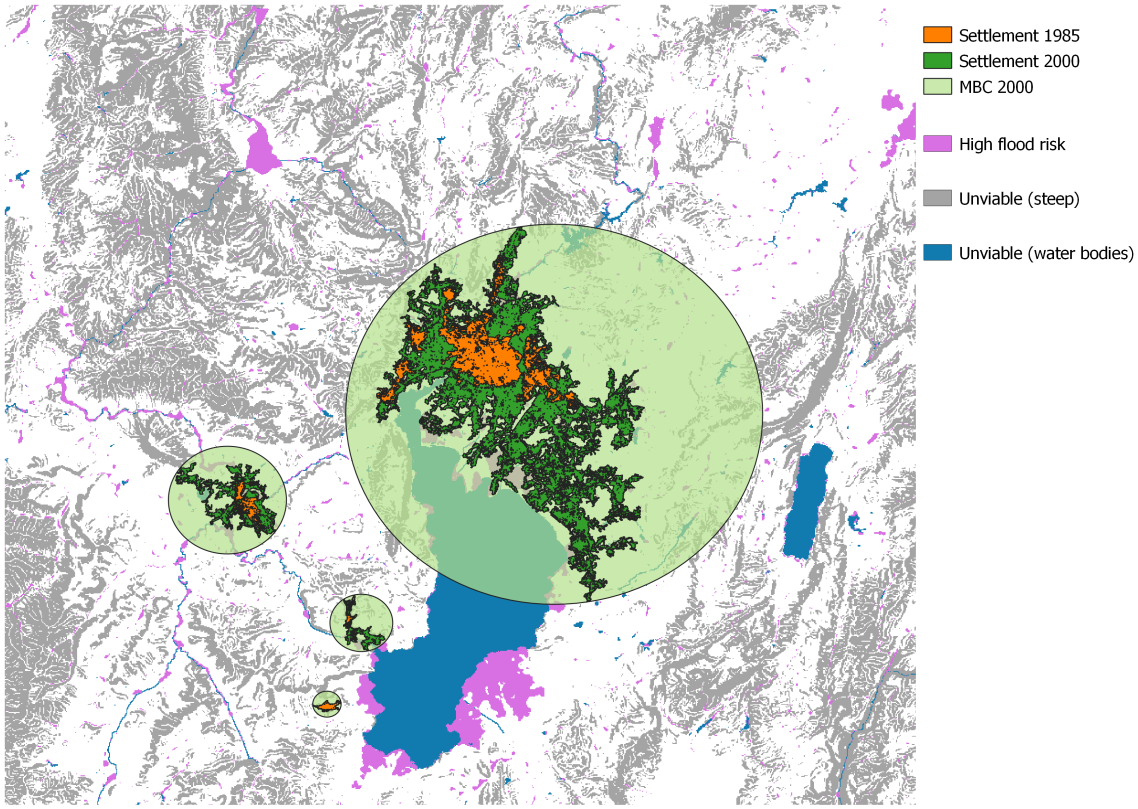
Before we seek to instrument the impact of land scarcity, we need to create a realistic measure of land scarcity around a city. To do this, we first need to delineate an area around a city that is a realistic measure of where the city could expand, and how constrained that land is. We name this area "Potential Developable Land" (PDL). We define a city's PDL by the land in the minimum bounding circle (MBC) around the city in the next period (in our setting in long differences, the next period is 15 years later). In Figure 4.4.1 and Figure 4.4.2, We show how we construct the PDL for 1985 using the MBC in 2000:

We put everything together in Figure 4.4.3: The PDL for settlements in 1985 (orange pixels) is the light green 2000 MBC (removing the built up orange settlements of 1985)[21]. In the same manner, the PDL for 2000 is the light red 2015 MBC removing the orange and dark green settlements of 1985 and 2000.

The MBC of the settlements at $t+1$ represents the option set open to urban planners to develop the city in the next period. This measure, if done correctly, should capture the endogenous development of the city towards water bodies or mountains.

[21]The PDL for 2000 is the MBC in 2015.

Figure 4.4.1: Minimum Bounding circles for settlement extents in 2000




We remove the city footprint at time $t$ within that minimum bounding circle as we consider developed land irreversible. We then calculate the share of land within that MBC that is unviable or in high-flood risk[22]. In our example, it is the share of pixels falling into the purple (high flood risk), grey (unviable steep) and blue (unviable water bodies) areas. In other words, we use future information of a city's actual expansion and backward induction to delineate what land was realistically considered during the 15 years of development. The PDL is a proxy for the complete set of choices considered by urban planners and policymakers for city expansion given its current location.

The change in land scarcity (our independent variable) is measured by computing the difference in how constrained city development was between the first 15-year period (1985-2000) and the second (2001-2015). We calculate the difference between the share of unviable land (steep terrain, oceans and rivers) and HFR land in the light red circles in 4.4.3 and the share of unviable and HFR land in the green circles, excluding the city footprint in 1985 and 2000

[22]As mentioned before, some cities merged over the period. In that case, we use the MBC encompassing all the continuous patches that will be merged by 2015

Figure 4.4.2: Resulting PDL in 1985: 2000 MBC (light green) with 1985 settlement extents cropped out

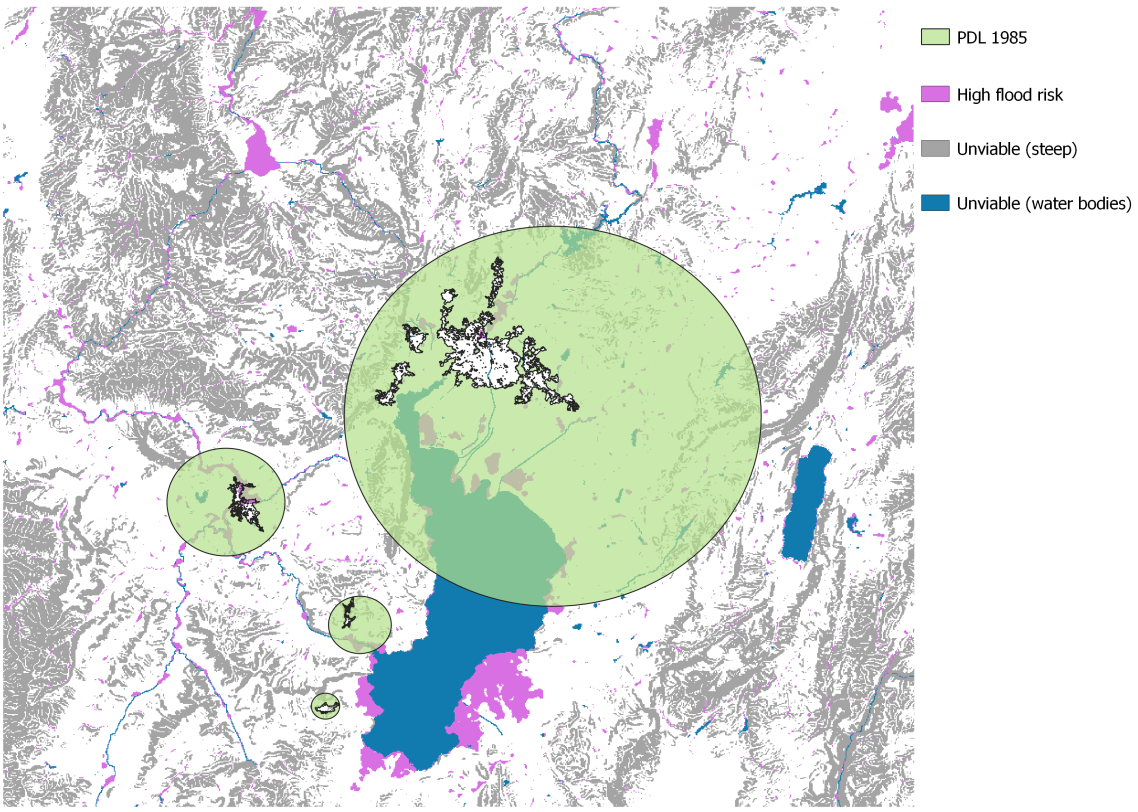
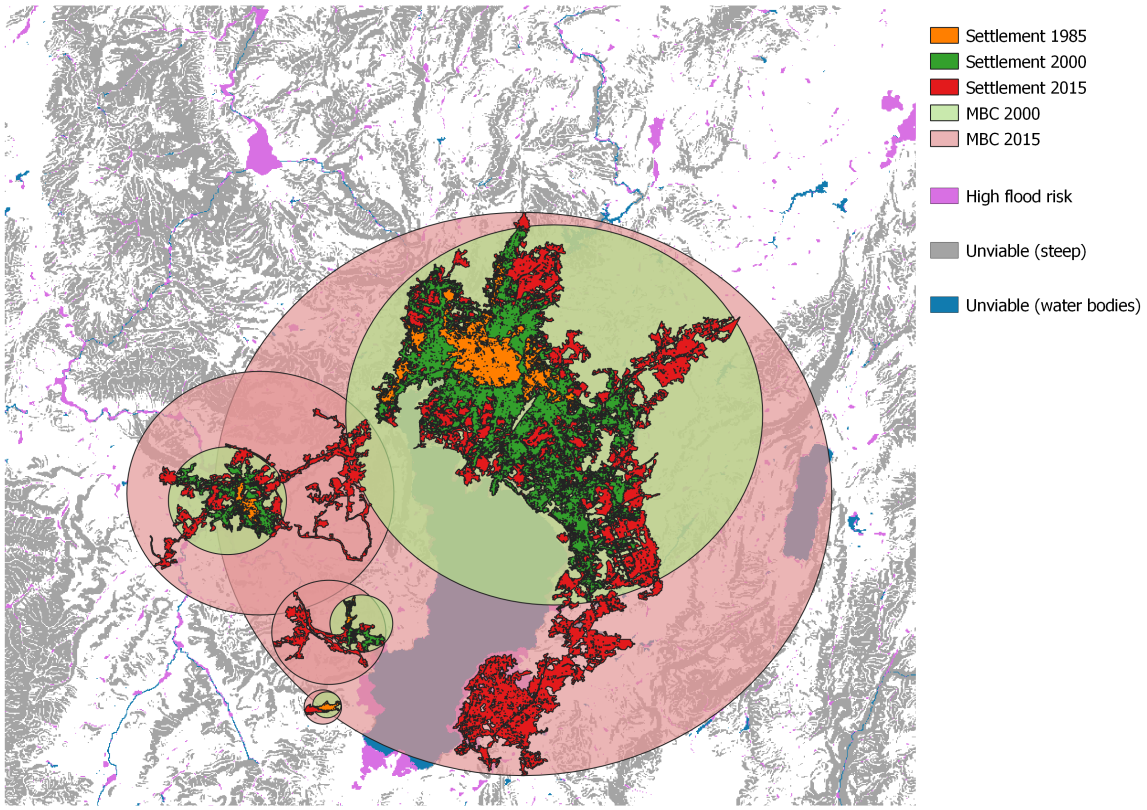


Figure 4.4.3: Settlement extents and minimum bounding circles




respectively. In Figure 4.4.3, we show 4 cities developing around a lake encircled by high flood risk areas. They are all distinct units of observation as they do not merge by 2015. The major city on the northeast side is developing toward the South between 1985 and 2000 (City 2000 in dark green). In the second period, the city is crawling on the East bank side of the lake (city 2015 in dark red). Notice that the share of unviable land and HFR areas increases between the green and light red circles as the city expands along the river. It is a realistic measure of endogenous growth towards unviable and HFR areas.

Now that we defined how we construct our independent variables, a naïve approach would run the cross-sectional regression of the share of new settlement extent (SE) in settlement i located in high flood risk (HFR), on the share of unviable land (steep land or water bodies) and the share of high flood risk land in the PDL for 1985 and 2000.

$$Y_{i,t} = \alpha_0 + \alpha_1 Share\ High\ Flood\ Risk\ in\ PDL_{i,t-1} + \alpha_2 Share\ unviable\ in\ PDL_{i,t-1} + \epsilon_{i,t} \quad (3)$$

with:

$$Y_{i,t} = (\frac{New\ SE\ HFR}{Total\ New\ SE})_{i,t} \tag{4}$$

$$Share\ High\ Flood\ Risk\ in\ PDL_{i,t-1} = (\frac{HFR\ area\ in\ PDL}{Total\ Viable\ area\ in\ PDL})_{i,t-1} \tag{5}$$

$$Share\ unviable\ in\ PDL_{i,t-1} = (\frac{Unviable\ land\ in\ PDL}{Total\ PDL\ area})_{i,t-1} \tag{6}$$

Now that we built a measure of change in actual land scarcity and HFR, we build on our main naive regression: we test whether a change in the 'bite' or share of unviable land and high flood risk land around a city leads to an increase in the share of settled land in high flood risk areas over the following 15-year period. We consider the timeframes of urban development to be large and split our panel dataset into two long time periods (1985-2000 and 2000-2015) to account for the length of time needed for cities to expand.

$$\Delta Y_{t,i} = \beta_0 + \beta_1 \Delta Share\ High\ Flood\ Risk\ in\ PDL_{i,t-1} + \beta_2 \Delta Share\ unviable\ in\ PDL_{i,t-1} + \Delta\epsilon_{i,t} \tag{7}$$

Given that:

$$\Delta Y_{t,i} = (\frac{New\ SE\ HFR}{Total\ New\ SE})_{i,2000-2015} - (\frac{New\ SE\ HFR}{Total\ New\ SE})_{1985-2000,i} \tag{8}$$

$$\Delta Share\ High\ Flood\ Risk\ in\ PDL_{i,t-1} = (\frac{HFR\ area\ in\ PDL}{Total\ Viable\ area\ in\ PDL})_{i,2000-2015} - (\frac{HFR\ area\ in\ PDL}{Total\ Viable\ area\ in\ PDL})_{1985-2000,i} \tag{9}$$

$$\Delta Share\ unviable\ in\ PDL_{i,t-1} = (\frac{Unviable\ land\ in\ PDL}{Total\ PDL\ area})_{i,2000-2015} - (\frac{Unviable\ land\ in\ PDL}{Total\ PDL\ area})_{1985-2000,i} \tag{10}$$

We show the results of this OLS estimation in Section 4.6. However, the main contribution

of the paper explores the bias of this regression. We worry that this regression will overestimate the impact of a change in land scarcity on settlement growth in high flood risk areas for two reasons.

We are faced with at least two sources of endogeneity due to the 'unobserved' quality of HFR land: (1) the original location of cities might be endogenous to unobserved amenity and productivity shifters linked with the land (e.g., close to a lake for easy to access to clean water) and (2) cities expand along the coast to derive more amenity and productivity value (in econometric terms, endogenous shifters). Our OLS estimate that we derive from the cross-sectional analysis above is a measure of the trade-off between risks and amenity value of developing around unviable or risky land, and it reflects the equilibrium outcome of an economic process.

The first source of endogeneity is well-documented. Cities locate originally near the coast, rivers or mountains for their productivity and amenity value. To deal with time-invariant sources of endogeneity, we take the long-difference of our outcome variable (the share of new settlements in HFR land) between our two time periods (1985-2000 and 2000-2015) to remove time-invariant factors such as the original location next to the coast or mountain. We thus investigate the effect of an increase in land scarcity (rather than the impact of scarcity in levels).

The second source of endogeneity is more subtle. Geographic barriers are seemingly exogenous. However, some economic forces (population growth, urban planning, amenity value of land) are driving cities to expand in high flood risk land and around mountains, constraining the amount of viable land around the city, therefore driving both our dependent and independent variables and introducing some omitted variable bias. Urban planners might indeed decide to expand along riverbanks, mountains or the coastline for the productivity and consumption value of that land. As a result, an increase in land scarcity and high flood risk land around a city is often the result of an endogenous decision to develop in the direction of, or near, unviable land. These effects could bias the role of land scarcity on settlement expansion in high flood risk areas. For example, if the coefficients $\beta_1$ and $\beta_2$ picked up the

voluntary development of a city towards scarce and risky land, it would overestimate the impact of increased land scarcity on settlement expansion in high flood risk areas. To solve this problem, we use an instrumental strategy to predict city growth that is not related to the directional city growth towards areas with unobserved productivity and amenity factors linked to flood risks.

In summary, the naïve regression set out above loads the effect of the endogenous growth of the city together with the amenity value of geographic barriers (rivers, oceans), solely on land scarcity. The main contribution of this paper is to construct an instrument for land scarcity that removes the endogeneity of voluntary city expansion towards high flood risk land due to unobserved amenity or productivity values of that land.

### 4.5 Instrumental Variable Construction

We base our instrument on Harari (2020). In her paper, Harari builds an instrument for city shape using the minimum bounding circle (MBC) of the historic maps for the city. She employs a mechanical model for city expansion to predict the area and shape that a city would take in a given year, based on its projected historical population growth. As explained in the previous section, we already use the MBC of the settlement extent in the next period as the definition for the current period PDL. However, this definition of the PDL reflects endogenous choices by city developers, who might choose to grow in high flood risk land for reasons other than land scarcity. We thus construct our instrumented PDL not as the MBC of the settlement extent in the next period (observed MBC for 2000 and 2015), but as the MBC of the settlement extent in the first period (1985) had it grown concentrically (predicted MBC for 2000 and 2015).

To do this, we construct the instrumented 'land scarcity' measure by first growing the city MBC in 1985 in all directions twice to obtain a predicted MBC for 2000 and 2015. We use an auto-regressive model with region and year fixed effects to avoid using potentially endogenous city growth rates. Using settlement extent growth rate to grow our instrument could lead to bias as more constrained cities could grow slower or faster (e.g. our theoretical model in Appendix Section 4.A.2)[23]. We run two robustness checks to test the implications of different

[23]Such that minimum bounding circles used in our IVs would be then larger and over-estimate the change in

city growth models.[24]

We then subtract MBC in the previous period (MBC 1985 and predicted MBC in 2000 respectively) to obtain a doughnut shape instrumented PDL.

Once we obtain that instrumented PDL for both periods, we calculate land scarcity and HFR land as we did for the PDL. We look at changes in land scarcity over time by calculating the difference in the share of unviable land and HFR land over the two time periods (the two doughnut shaped instrumented PDLs). The long differences allow us to remove the endogeneity of the initial city location, so we recover exogenous variation in land scarcity. The main assumption behind our instrument is that the relative distance between geographic obstacles as the city expands is exogenous. In this regard, this instrument can be thought of as a differentiated spatial Bartik instrument.

To summarise, we recover random changes in land scarcity around the city by exploiting the changes in unviable and HFR land captured by the random location of geographic barriers around a city as a city expands following a predictive city growth model. We document the construction of the instrument visually below.

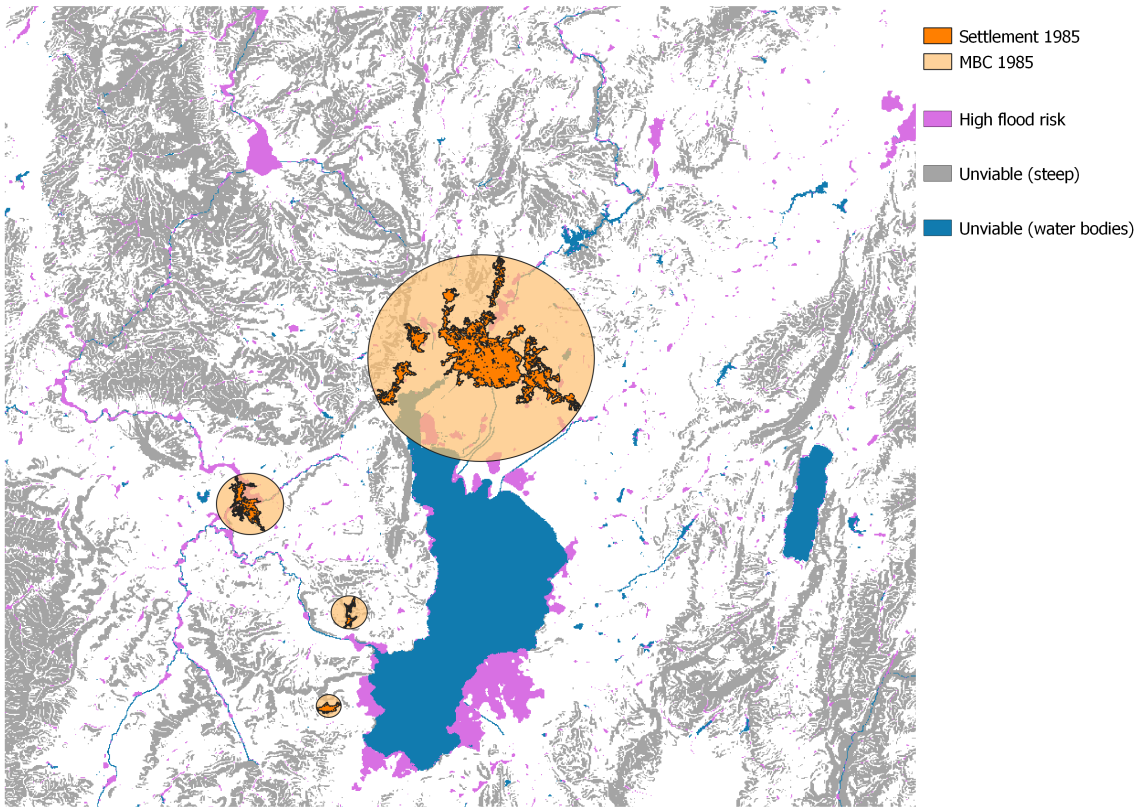
The original settlement extent MBC in 1985 shown in Figure 4.5.1.

We predict settlement extent expansion using an autoregressive model with region and period fixed effects. The instrumented 2000 MBC is shown in Figure 4.5.2 (circle in light green). We replicate this step for 2015: we grow the 1985 MBC again using the predicted radius for 2015 and remove the area of the instrumented 2000 MBC. The resulting instrumented 2015

scarcity and HFR.

[24]We also build a second version of the instrument, where we impose that the area of the instrumented PDL is of equal size to the actual PDL for each time period (Appendix 4.A.3). Our results are not sensitive to this specification, lending credence to our instrument. We conduct a second robustness check where we compare the actual city growth rate obtained using the auto-regressive regional growth model with growth rates systematically smaller or larger. To create the systematically smaller and larger growth rates, we pick for each city a shifter between 5% and 20% randomly generated from a uniform distribution. We find that smaller growth rates have smaller changes in unviable land but slightly larger changes in HFR land. The difference are, however, very small (Tables 4.A.4 amd 4.A.5). As a result of this, in order to minimise measurement error and bias introduced across cities, we prefer using the regional growth model instead of fixing the area of the actual and instrumented PDL.

Figure 4.5.1: Original 1985 Minimum Bounding Circle

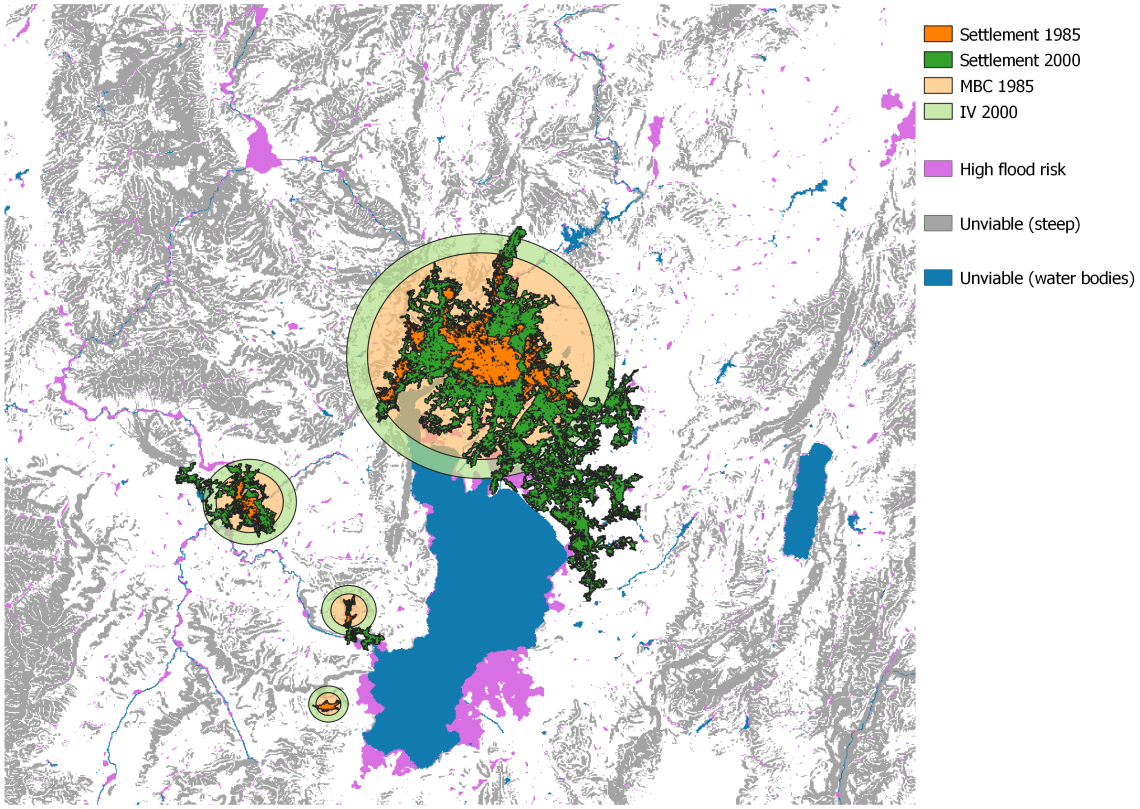



*Step1: Create the original Minimum Bounding Circle for the 1985 City*

MBC is shown in Figure 4.5.3 (circle in light red). We thus obtain predicted settlement extents in 2000 and 2015. From these predicted areas, we can calculate their predicted radius. When we grow the 1985 MBC using the predicted radius for 2000, we remove the original 1985 MBC shape.

As mentioned previously, the predicted MBC for 2000 and 2015 in light green and light red in Figures 4.5.2 and 4.5.3 are centred on the original 1985 MBC centre[25]. In the last step, we estimate the difference in the share of unviable land and the share of HFR areas between the red and green instrumented MBCs.

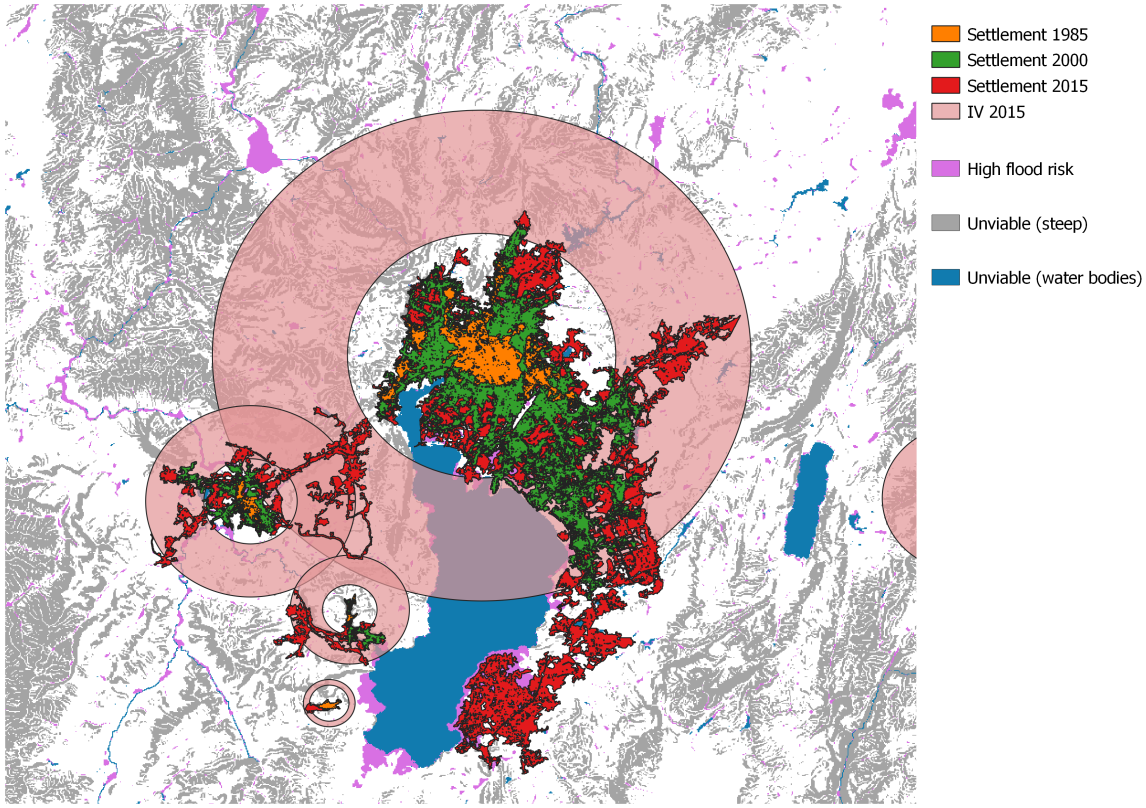
Figure 4.5.2: Instrumented 2000 MBC




*Step 2: Grow the 1985 MBC using predicted radius from an autoregressive model with region and period fixed effect; remove the original 1985 MBC shape*

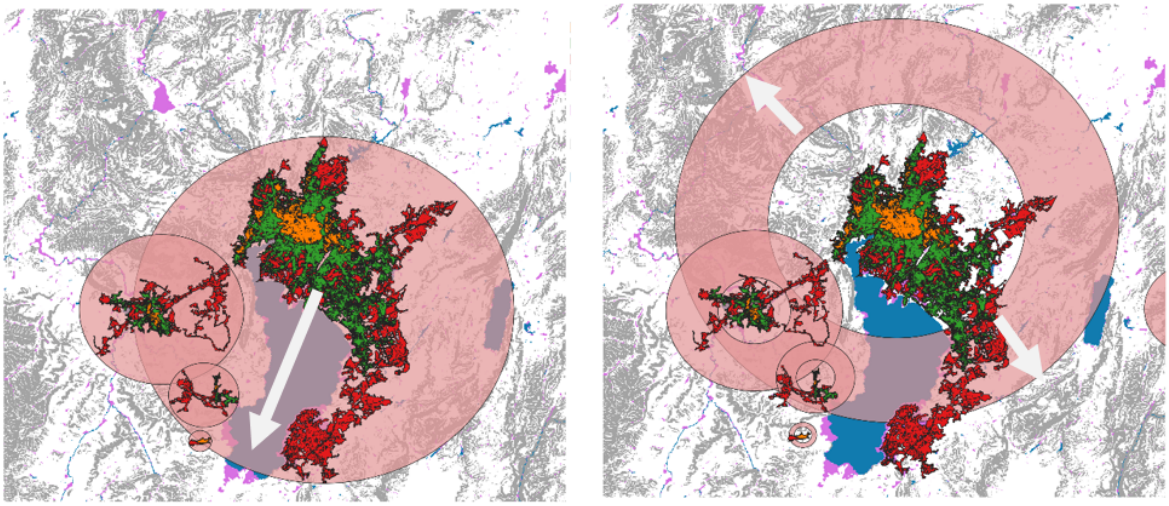
[25]Ideally, we would use historic city centres, but we do not have this information globally. The strong underlying assumption is that the original MBC centres in 1985 coincide with the historic city centres.

Figure 4.5.3: Instrumented 2015 MBC




*Step 3: Grow again the 1985 MBC using predicted radius from the same autoregressive model with region and period fixed effect; remove the instrumented 2000 MBC*

Figure 4.5.4: PDL versus instrument: a visual explanation




*The PDL is capturing the directional growth towards the lake. The instrument is capturing the land around the city in all directions*

The main difference between the naive way to capture land scarcity (using the PDL, defined as the MBC of the city in the next period minus the city shape) and the instrument (using the original MBC that we grow concentrically using the regional auto-regressive model) is the directional growth of the city. To explain the difference visually, we use our example city in Figure 4.5.4. Between 1985 and 2000, the PDL centroids shift towards the lake as the city grows. In the instrument, the MBC buffers stay centred around the original centre of the city. As a result, the PDL area captures more and more of the unviable and less and less of the available land in the north.

If our cities were perfectly monocentric and growing in all directions, our two measures would be similar. As with Harari (2020), variation in land scarcity here does not stem from the sole presence of water bodies or steep slopes, nor from their size, but from the relative position in space of these constraints. While the relative position of these constraints might be correlated to the source of endogeneity that we worry about, the changes in the relative position of these constraints over time does not. In other words, our difference in difference estimation is crucial in our estimation strategy.

The instrumental variable, on top of the long difference specification, removes the bias because the variation in the share of unviable land and HFR land around a city relies on the changes in unviable land captured by geography interacted with a mechanically predicted model of city growth. This interaction excludes, by construction, the variation resulting from endogenous policy choices or from unobserved characteristics (amenity, productivity, policy choices). In other words, the instrument estimates the 'mechanical' role of land scarcity net of the amenity value, compared to the amenity-based role of land scarcity captured by the OLS.

Our instrument also captures the fact that safe land was available around the city despite the city not developing there – hence the risky growth was not the result of unviable land but rather due to the choice of cities to develop in that direction. In this case, the city could have developed on the Northeast side instead of the lake banks.

Formally, we want to instrument the $\Delta$*Share High Flood Risk in PDL*$_{i,t-1}$ and $\Delta$*Share unviable in PDL*$_{i,t-1}$ in our PDL with the same measure in the predicted IV MBCs, both in 2000 and 2015.

$$\Delta \textit{Share High Flood Risk in IV}_{i,t-1} = \left(\frac{\textit{HFR area in IV}}{\textit{Total Viable area in IV}}\right)_{i,2000-2015} - \left(\frac{\textit{HFR area in IV}}{\textit{Total Viable area in IV}}\right)_{1985-2000,i} \tag{11}$$

$$\Delta \textit{Share unviable in IV}_{i,t-1} = \left(\frac{\textit{Unviable land in IV}}{\textit{Total IV area}}\right)_{i,2000-2015} - \left(\frac{\textit{Unviable land in IV}}{\textit{Total IV area}}\right)_{1985-2000,i} \tag{12}$$

Our instrument removes endogeneity in a long-difference specification. Recall that we consider the timeframes of urban development to be large and look at long differences between two time periods (1985-2000 and 2000-2015) to account for the length of time needed for cities to expand. We test whether a change in how binding unviable land and high flood risk land are around a city, leads to an increasing share of new city expansion in high flood risk areas over the following 15-year period. Our main dependent variable ($Y_{t,i}$) is the difference in the share of new settlement extent (SE) in a city $i$ that is at high flood risk (HFR) between two 15-year periods (2000-2015 and 1985-2000).

Our main estimation strategy thus boils down to:

$$\Delta Y_{t,i} = \beta_0' + \beta_1' \Delta \textit{Share High } \widehat{\textit{Flood Risk}} \textit{ in PDL}_{i,t-1} + \beta_2' \Delta \textit{Share } \widehat{\textit{unviable in}} \textit{ PDL}_{i,t-1} + \Delta \epsilon_{i,t}' \tag{13}$$

Where the independent variables are the predicted values that result from instrumenting the change in the "naive" measure of the share of unviable land and high flood risk land (through the potential developable land metric - PDL) with the change in the share of unviable land and high flood risk land using our instrument (the predictive model of city shape with re-centred MBCs).

## 4.6 Results

### 4.6.1 Impacts on Settlement Growth in High Flood Risk Areas

As set out in Section 4.4, our paper is not about one estimate of the role of high flood risk land and unviable land on risky urbanisation. Rather, the contribution of our paper is the comparison of three estimates to parse out the different ways land scarcity and unsafe land drive cities to expand into high flood risk areas.

We indeed compare three estimates, as we move from (1) a cross-sectional regression that captures all three aspects of land scarcity: the geographic barriers themselves, the initial city location and the amenity value of the constraints; to (2) a long-difference regression using the potential developable land (PDL) metric, which captures both the role of increased land scarcity due to geographic barriers and of the directional amenity-based city growth towards those barriers; to (3) an instrumental variable approach in long differences which captures the role of the change in land scarcity only, net of the amenity value of geographic constraints and net of the amenity-based directional growth of the city.

Our results cover all of China, which consists of 2555 settlements in our dataset. China is responsible for almost 50% of high-risk settlement growth over our study period. For each flood risk, we only keep cities that are exposed to flood risks in their PDL during the period (1985-2015). It leaves us with 2036 settlements facing any flood risk.

Table 4.6.1 presents the results of the first naïve OLS cross-sectional regressions. Given their separate amenity value, we control for the three different types of unviable land separately: steep land (a slope superior to 15%[26]), rivers and oceans. The share of high-flood risk pixels and unviable pixels explains a large part of the variation in new settlements in high-flood risks in the two periods. The relationship is almost one-to-one for high flood risk. The main source of unviable land is the presence of steep land. An increase in one percentage point share of surrounding steep land is associated with a 0.25 percentage point increase in new settlements in HFR areas. The presence of rivers seems to increase new SE in HFR in the second period

[26] Our results are not sensitive to the degree of the slope

but the association is much weaker than it is for steep land. Oceans do not seem to be a factor – however as we will expand on later, this might reflect the fact that coastal cities tend not to be monocentric and thus our PDL does not capture the presence of unviable land and HFR areas equally well for these cities. These correlations reflect not only the impact of the geographic barriers themselves but also the original placement of cities and the amenity-based growth of cities towards high flood risk areas.

Table 4.6.1: New share SE in HFR in cross-sectional OLS (Any flood risk)

| | (1) | (2) |
|---|---|---|
| | 1985 | 2000 |
| Share HFR | 0.933** | 0.990** |
| | (0.0470) | (0.0332) |
| Share unviable (steep) | 0.270** | 0.245** |
| | (0.0467) | (0.0361) |
| Share unviable (rivers) | 0.0170 | 0.130* |
| | (0.0514) | (0.0557) |
| Share unviable (ocean) | -0.0409 | 0.0192 |
| | (0.0795) | (0.0674) |
| N | 2036 | 2036 |
| Rsquared | 0.765 | 0.787 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Table 4.6.2 presents our first stage results of the four instruments: share of high flood risk land around a city, share unviable steep land around a city, share of rivers around a city, and share of oceans around a city.

Our instruments are all significant predictors at the 1% for their related outcomes. The F-statistics of Columns (1),(2) and (3) are significantly higher than 10. The F-statistic for the share of unviable land coming from oceans is lower than 10. This can be explained by two data limitations: we only have a limited number of coastal cities (about 200), and they tend to merge together more often than other cities, creating long urban patches along the coast. Our PDL measure and instrument might not be the most appropriate for these cities, and the predictive power of the instrument will be lower. We urge caution when interpreting the

Table 4.6.2: First stage (Any flood risk)

| | (1) Δ Share HFR (any risks) | (2) Δ Share unviable (steep) | (3) Δ Share unviable (rivers) | (4) Δ Share unviable (ocean) |
|---|---|---|---|---|
| Δ Share HFR IV (any risk) | 0.297** | -0.00788 | 0.0219* | -0.00705 |
| | (0.0354) | (0.0116) | (0.00820) | (0.00900) |
| Δ Share unviable IV (steep) | 0.00270 | 0.292** | 0.00541 | -0.0147 |
| | (0.0148) | (0.0320) | (0.00855) | (0.00849) |
| Δ Share unviable IV (rivers) | -0.00358 | -0.0389** | 0.174** | 0.0128 |
| | (0.0177) | (0.0132) | (0.0195) | (0.00865) |
| Δ Share unviable IV (ocean) | 0.134** | -0.0194 | 0.0316** | 0.335** |
| | (0.0439) | (0.0308) | (0.00952) | (0.109) |
| N | 2036 | 2036 | 2036 | 2036 |
| F | 21.30 | 29.22 | 26.91 | 7.657 |

*$p < .05$, **$p < .01$
Standard errors clustered at the region level

results for coastal cities. In order to avoid cross-predictions in our instruments, when looking at flood risk separately, we only include steep land for pluvial risk and steep land and rivers for fluvial risks.

Table 4.6.3 reports estimates from the long difference ordinary least squares (OLS) estimation and the long difference instrumental variable (IV) for our main outcome: the share of new settlement of a city $i$ in high flood risk areas.

Column 1 in Table 4.6.3 presents results from an OLS specification in long-difference. The OLS results provide estimates of the intention to treat, and the IV results provide estimates of the average treatment effect on the treated. The OLS estimate tells us that an increase in the share of high flood risk land around a city between 1985 and 2000 leads cities to expand more in high flood risk land between 2000 and 2015. Compared to the cross-sectional results, the coefficient is smaller than in a one-to-one relationship. About a third of the variation is captured by city-invariant characteristics. Cities are actually managing to build away from high flood risk land even when they are increasingly encircled by it.

Once we instrument for the share of high flood risk land using the growing monocentric

Table 4.6.3: New Share SE in HFR - Long-difference and IV (Any flood risk)

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (any risks) | 0.605** | 0.540** | 0.615** | 0.632** |
| | (0.0727) | (0.191) | (0.0651) | (0.204) |
| Δ Share unviable (steep) | | | 0.0397 | 0.0390 |
| | | | (0.0666) | (0.194) |
| Δ Share unviable (rivers) | | | -0.105 | -0.577 |
| | | | (0.116) | (0.361) |
| Δ Share unviable (ocean) | | | 0.467* | 0.870 |
| | | | (0.177) | (0.450) |
| N | 2036 | 2036 | 2036 | 2036 |
| Rsquared | 0.118 | 0.117 | 0.129 | 0.105 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

circles (Column 2), we see a small but not significant decrease in the role of neighbouring HFR land on new SE in HFR. There is path dependence of cities expanding into high flood risk land when they are near areas with high flood risk to start with, even when we instrument for the endogeneity of nearby high flood risk land. The coefficients for high flood risk land remain significant throughout our four specifications, even when we control for the share of unviable land. This is new evidence for the path dependence story: cities are locked in to a path of city expansion into high flood risk land.

Turning to Columns 3 and 4 where we control for different forms of land scarcity due to geographic barriers, we find that an increase in the share of unviable land is no longer significant aside from the unviable land from the ocean. Cities growing in HFR areas are originally located near rivers and steep land, but once we remove the time-invariant characteristics, there is no evidence that an increase in unviable land is a factor in HFR growth. The coefficients from the IV are also not significant. As mentioned above, our PDL and IV are weak for coastal cities. Introducing a weak instrument might be biasing our estimation. We thus do the analysis by flood risk type separately.

### 4.6.2 Heterogeneity by flood risk type

We proceed to study heterogeneity by flood risk type. The most common flood risks in our sample are fluvial floods, caused by overflowing rivers. The second most common is pluvial floods, where intensive rain causes excess water, which cannot be absorbed fast enough by the drainage systems or by ground infiltration. Pluvial risks are particularly problematic in urban areas as they are less predictable. While pluvial floods occur often, the amount of land subject to pluvial floods is small. The least common type of flooding is coastal. Due to high tides and storms, it is particularly dangerous as water levels can reach higher levels of inundation depth than other types of floods.

We first present our estimation results for fluvial risks only in Table 4.6.4. We include the first stages of the results in the Appendix (1.7.3.3). Most cities in our sample (1771 out of 2555) are exposed to fluvial risks – the results are thus fairly similar to the general flood risk analysis presented above. In Column 1, we find that once we take into account the original location of the city, the impact of an increase in flood risks in the PDL is halved. Similarly, in Column 3, the positive relationship between scarcity and HFR settlement growth disappears.

Table 4.6.4: New Share SE in HFR - Long-difference and IV (Fluvial flood risk)

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (fluvial) | 0.509** | 0.393* | 0.524** | 0.513** |
| | (0.0639) | (0.154) | (0.0617) | (0.170) |
| Δ Share unviable (steep) | | | 0.0487 | 0.0573 |
| | | | (0.0579) | (0.159) |
| Δ Share unviable (rivers) | | | -0.0760 | -0.706 |
| | | | (0.0900) | (0.352) |
| N | 1771 | 1771 | 1771 | 1771 |
| Rsquared | 0.0909 | 0.0862 | 0.0921 | 0.0518 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

In Columns 1 and 2, we see that the coefficients drop from 0.5 to 0.39, yet the difference is not statistically significant. The coefficients on scarcity are still null, with some indications

that they might even become negative in HFR areas. This might be explained if higher shares of unviable land decrease city growth overall. We do not observe density in our results – we instead focus on the extensive growth - but it might be an indication that constrained cities become denser.

Turning to pluvial risks in Table 4.6.5, we find in Columns 1 and 3 that the role of HFR areas is again almost halved compared to the cross-sectional specification. In other words, cities manage to develop away from HFR. In Columns 2 and 4, we also see that once instrumented for, the role of HFR land also becomes non-significant and negative, which could once again be due to cities expanding less when faced with high surrounding flood risk. We also see that once instrumented for, the role of HFR land also becomes non-significant (between columns 3 and 4). One hypothesis for this result is that pluvial land covers, on average, a far smaller share of the potential developable land of a city, so pluvial flood risk is less 'binding': cities are more likely to successfully develop away from that land.

Table 4.6.5: New Share SE in HFR - Long-difference and IV (Pluvial flood risk)

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (pluvial) | 0.592* | -0.837 | 0.613* | -0.681 |
| | (0.284) | (0.590) | (0.280) | (0.547) |
| Δ Share unviable (steep) | | | 0.0384 | 0.115 |
| | | | (0.0230) | (0.0842) |
| Δ Share unviable (rivers) | | | -0.0595* | -0.0722 |
| | | | (0.0226) | (0.0772) |
| N | 1732 | 1732 | 1732 | 1732 |
| Rsquared | 0.0336 | . | 0.0383 | . |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Finally, we include the results of coastal flood risk in Table 4.6.6. We do not find any statistically significant results after taking the long difference. While it could mean that most of the city growth among coastal cities is endogenous to amenities and to the productivity value of the coast – if coastal cities could develop monocentrically inland instead of growing

along the coast – it could also mean that our framework is likely not suited to coastal growth. Coastal cities tend to have merged with other coastal cities over our time period, creating units of analysis that are not well captured by our framework (both by the potential developable land metric and the instrumented land scarcity metric).

Table 4.6.6: New Share SE in HFR - Long-difference and IV (Coastal flood risk)

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (coastal) | 0.159 | -0.588 | 0.165 | -0.920 |
| | (0.129) | (0.539) | (0.125) | (1.092) |
| Δ Share unviable (steep) | | | 0.0174 | 3.993 |
| | | | (0.131) | (6.437) |
| Δ Share unviable (ocean) | | | -0.0878 | 0.570 |
| | | | (0.110) | (0.864) |
| N | 190 | 190 | 190 | 190 |
| Rsquared | 0.0121 | . | 0.0142 | . |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Our results show that the narrative that scarcity of land is forcing cities to expand into high flood risk land does not hold. Naïve regressions would overestimate the role of unviable land in driving urbanisation in high flood risk land, when it seems as if all types of geographic barriers (steep land, rivers and oceans) are not binding the city to grow towards risky land. Urban growth in risky flood zones is likely driven by initial location and some consumption or amenity value of that land leading cities to continue expanding towards it (e.g., coastal amenities).

## 4.7 Conclusion

In this paper, we study the expansion of cities in high flood risk areas in China. We show that the increasing expansion of cities in high flood risk areas is largely due to a lock-in effect of cities surrounded by high flood risk. No matter which direction they expand in, they will face some flood risk and cannot escape building in this area.

We provide causal evidence that, once instrumented for, "hard constraints" such as mountains or water bodies are not a driver of urbanisation in high flood risk areas. Instead, risky growth seems driven by the deliberate consumption of flood-risk land due to the amenity value associated with the geographic barriers (e.g. rivers).

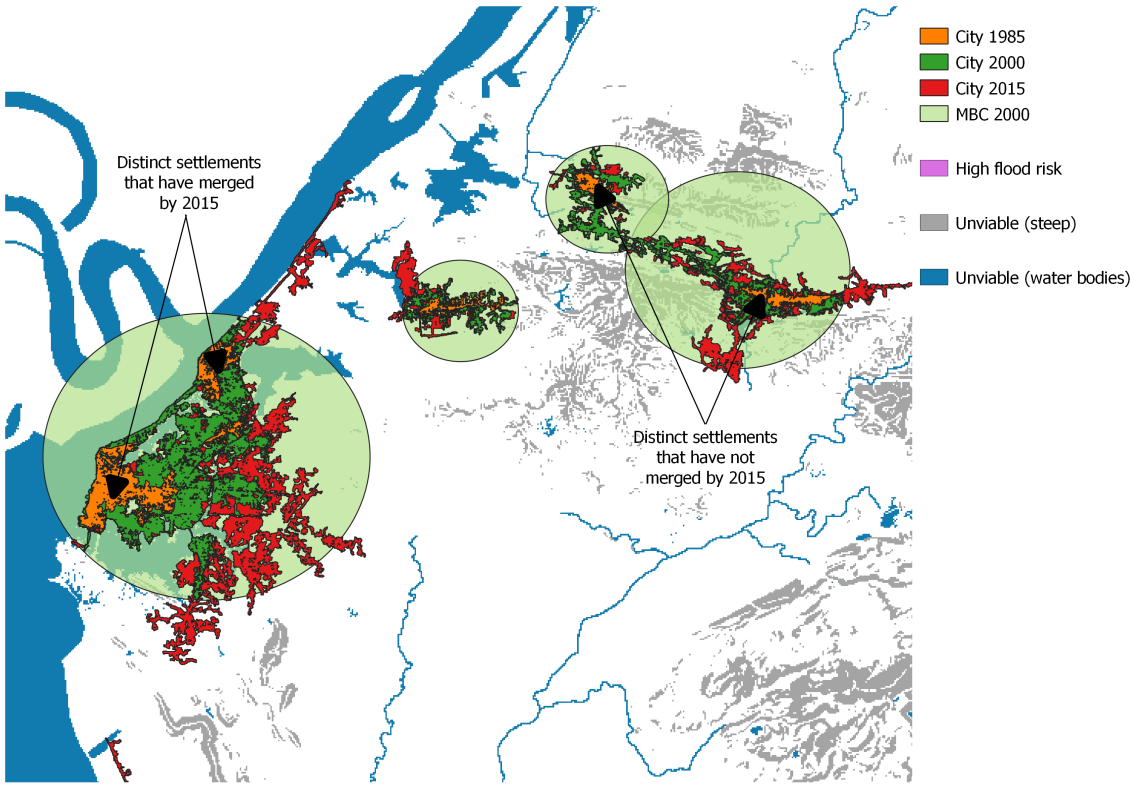
Our results bring new evidence that cities are locked in to a path of risky growth and that the expansion of cities in high flood-risk areas will persist. The future damages of floods, accounting for heightened flood risks due to climate change and sea-level rise, will continue to grow, warranting large investments in adaptation technologies such as drainage systems or dikes.

It is not in the scope of this paper to comment on the role of stricter urban planning as it would require accounting for other factors such as commuting costs, rents and wages. A promising extension of this work would introduce land scarcity from urban planning into a computational general equilibrium model. Future research should also seek to study whether land scarcity plays a role for the entire distribution of flood hazards (e.g. inundation depth and return rates), as well as combine data on density to capture not only the expansion of cities but also the vertical growth of cities in high flood risk zones.

## Appendix

### 4.A.1 Additional Maps

Figure 4.A.1: Example of settlements that have merged into one continuous urban patch by 2015




*Note: The first settlement on the left is composed of several distinct settlement patches in 1985 (orange polygons). As they merged into one continuous patch in 2000 (green polygon) and then expanded further out in 2015. On the other hand, the two patches on the right side haven't merged yet and thus are counted as separate units. For each unit of analysis, the PDL is defined as the Minimum Bounding Circle in the next period. As the settlements on the left have merged, they have one PDL. On the contrary, the settlements on the right have two distinct PDLs.*

### 4.A.2 Simplified Monocentric City Model Framework with Floodable and Unviable Land

We want to study whether settlements bound by unviable land have a greater share of development in risky areas. Recall that our definition of land scarcity includes both unviable

land (geographic barriers: steep land, rivers and oceans) and unsafe land (high flood risk land). We develop a simple monocentric city model with land scarcity and flood risk to examine the impact of land scarcity.

In a classic monocentric city model, employment is centralized in the central business district. The model considers a city that is small (one of many in a nation) and open, where people move costlessly between cities. For simplicity, we assume equal utility in the city $\bar{u}$, rent $R(x)$ is a function of commuting costs $tx$ as a linear function of distance to the CBD, wages $w$ are equal everywhere, and lot size is equal to 1 such that population density is the same. To connect the model to our empirical investigation, note that the key parameters of interest are the total city population size $N$, and the edges of the city ($x^{max}$). It gives us the basic bid rent function:

$$R(x) + 2tx = 0 + 2tx^{max}$$

$$R(x) = 2t(x^{max} - x)$$

then add two dimensions to the model: floodable land $F$ on one side of the city, and unviable land $d$ on the other side, which can be thought of as a mountain or a water body (so segment $d$ is unoccupied).

Floodable land adds two variables to the utility function: dis-amenity $A$ of flood risk and amenity value $B$ of proximity of the coast or river. People located in a floodable area $F$ suffer from the dis-amenity cost $A$, where $A < 0$.

As our model compares the impact of flood risk between cities, we simplify the amenity value $B$ of the coast or the river to be equally distributed throughout the city, rather than as a function of the distance to the coast. Wages $w$ are unchanged. The line gradient is unchanged as the amenity and dis-amenity parameters do not change as a function of distance.

The size and total population of the city are affected by the two new variables $A$ and $B$.

First, we focus on segment $F$:

$$R(x \in F) + 2tx + A + B = 0 + 2tx^{max} + B$$

$$0 \leq R(x \in d) = 2t(x^{max} - x) - A$$

Whether the segment $F$ is inhabited depends on the value of $A$, $t$ and the edge of the city $x^{max}$. We are left with three cases; high flood risk land is occupied up to $x_b \in [x_W, x'_W]$ as a function of $A$:

1. $A$ is big: $d$ is unoccupied:
$$2t(x'^{max} - x) - A \leq 0$$
2. $A$ is small: $d$ is occupied
$$2t(x'^{max} - x) - A > 0$$
3. $A$ is intermediate: the area $x_W$ to $x_b$ is occupied and the area $x_b$ to $x'_W$ unoccupied,
$$2t(x'^{max} - x_b) - A = 0$$

The total size of the city is thus lower by a factor of $A/2t$ as $A > 0$. An intuition for $A/2t$ is the dis-amenity of the floodable land standardised by the commuting distance of that land from the CBD.

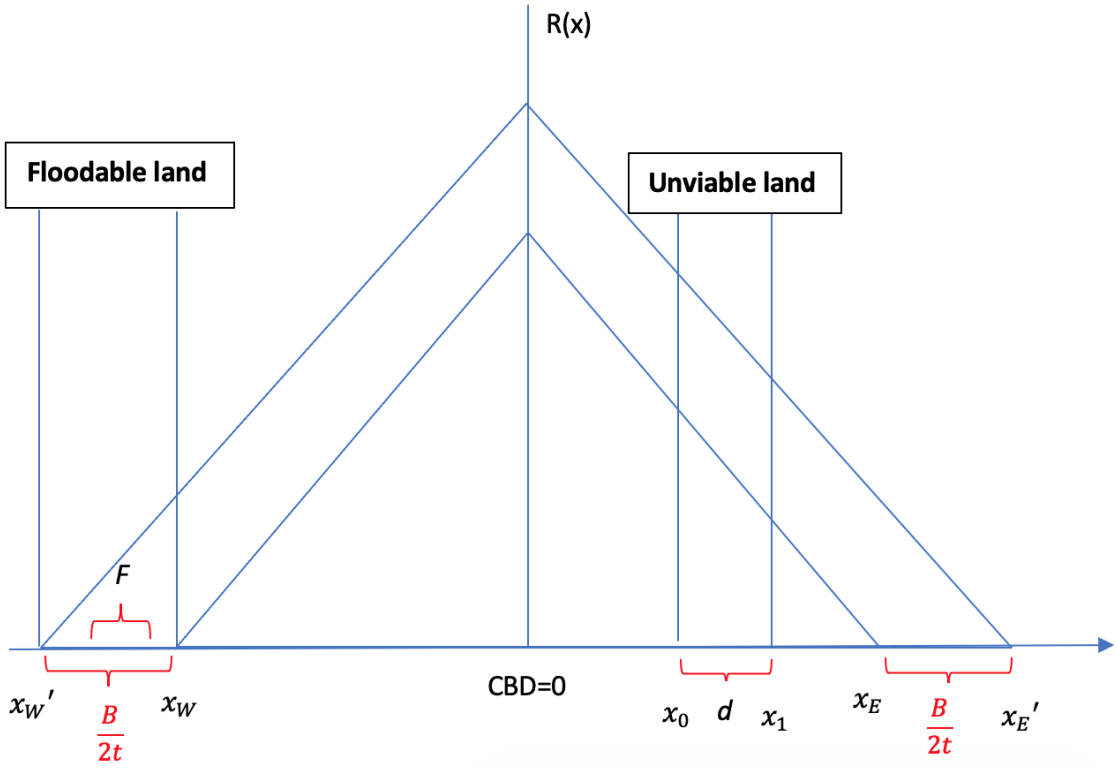
Then we look at the total population in the city given by population on both sides of the city:

$$N' = N'_W + N'_E = x'_W + x'_E - d$$

Utility at the CBD is given by:

$$w - R(0) - 2tx'^{max} - B = \bar{u}$$

Figure 4.A.2: Monocentric City Model Framework with Floodable and Unviable Land




Putting everything together, the population on the east side and west side is now:

$$N'_E = x'_E - d = (w - \bar{u} + B)/2t - d$$

$$N'_W = (w - \bar{u} + B)/2t + A/2t$$

Here $x'_E$ and $x'_W$ are the edges of the new city, after the dis-amenity and amenity shock $A$ and $B$. $N'_E$ and $N'_W$ correspond to the population after the shock on the East and the West sides of the city, which are expanded by a factor $B/2t$, the amenity value standardized by commuting cost. The amenity $B$ ($B > 0$, it is a positive amenity) has an impact on the equilibrium population.

As mentioned previously, the population in $N_F$ is lower by the factor $A/2t$ as $A > 0$. The total city population is:

$$N'_E = x'_E - d = \frac{w - \overline{u} + B}{2t} - d$$

$$N'_W = \frac{w - \overline{u} + B}{2t} + \frac{A}{2t}$$

$$N' = N'_E + N'_W = \frac{w - \overline{u} + B}{t} + \frac{A}{2t} - d$$

The city will also move further away from the monocentric model as a function of the size and location of unviable land. We show that the population will settle in high flood risk land if they value the amenity value of the land over the dis-amenity value of the flood, given the commuting benefits associated with the land.

### 4.A.3 Additional Results

**4.A.3.1 Robustness Checks** We first build a second version of the instrument, where we impose that the area of the instrumented PDL to be the size of the actual PDL for each time period. Our results are not sensitive to this specification, which lends credence to our instrument.

Table 4.A.1: New Share SE in HFR - Long-difference OLS and IV (Any flood risk) - IV area equals PDL area

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (any risks) | 0.631** | 0.490* | 0.641** | 0.488* |
| | (0.0728) | (0.184) | (0.0661) | (0.197) |
| Δ Share unviable (steep) | | | 0.0405 | -0.0152 |
| | | | (0.0662) | (0.165) |
| Δ Share unviable (rivers) | | | -0.129 | -0.172 |
| | | | (0.122) | (0.210) |
| Δ Share unviable (ocean) | | | 0.455** | 0.729 |
| | | | (0.164) | (0.401) |
| N | 2047 | 2047 | 2047 | 2047 |
| Rsquared | 0.129 | 0.123 | 0.141 | 0.129 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Table 4.A.2: New Share SE in HFR - Long-difference OLS and IV (Fluvial flood risk) - IV area equals PDL area

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (fluvial) | 0.509** | 0.389* | 0.523** | 0.438* |
| | (0.0637) | (0.177) | (0.0615) | (0.170) |
| Δ Share unviable (steep) | | | 0.0499 | 0.107 |
| | | | (0.0573) | (0.130) |
| Δ Share unviable (rivers) | | | -0.0721 | -0.267 |
| | | | (0.0891) | (0.195) |
| N | 1781 | 1781 | 1781 | 1781 |
| Rsquared | 0.0909 | 0.0859 | 0.0921 | 0.0826 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Table 4.A.3: New Share SE in HFR - Long-difference OLS and IV (Pluvial flood risk) - IV area equals PDL area

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| Δ Share HFR (pluvial) | 0.592* | -0.942 | 0.613* | -0.895 |
| | (0.283) | (0.773) | (0.280) | (0.776) |
| Δ Share unviable (steep) | | | 0.0385 | 0.0603 |
| | | | (0.0230) | (0.0656) |
| Δ Share unviable (rivers) | | | -0.0598* | -0.159* |
| | | | (0.0226) | (0.0637) |
| N | 1742 | 1742 | 1742 | 1742 |
| Rsquared | 0.0336 | . | 0.0383 | . |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

We conduct a second robustness check where we compare the actual city growth rate obtained using the auto-regressive regional growth model with growth rates systematically smaller or larger. To create the systematically smaller and larger growth rates, we pick for each city a shifter between 5% and 20% randomly generated from a uniform distribution. We find that smaller growth rates have smaller changes in unviable land but slightly larger changes in HFR land. The differences are, however, very small (Tables 4.A.4 and 4.A.5).

Table 4.A.4: Δ Share non-viable by IV buffer size

| | (1) Any non-viable | (2) Ocean | (3) Rivers | (4) Steep |
|---|---|---|---|---|
| Larger MBC | 0.00162** | 0.000303** | 0.000110 | 0.00125** |
| | (0.000283) | (0.0000972) | (0.000168) | (0.000222) |
| Smaller MBC | -0.00162** | -0.000295** | -0.0000588 | -0.00133** |
| | (0.000295) | (0.0000877) | (0.000186) | (0.000229) |
| N | 7662 | 7662 | 7662 | 7662 |
| Rsquared | 0.000492 | 0.000135 | 0.00000338 | 0.000493 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the city level
The base category is actual buffer size calculted using regional growth

Table 4.A.5: Δ Share HFR by IV buffer size

| | (1) Any HFR | (2) Fluvial | (3) Pluvial | (4) Coastal |
|---|---|---|---|---|
| Larger MBC | -0.000777** | -0.000578** | 0.000124 | -0.0000114 |
| | (0.000286) | (0.000219) | (0.0000816) | (0.000405) |
| Smaller MBC | 0.000724* | 0.000670** | -0.000142 | 0.00000866 |
| | (0.000298) | (0.000246) | (0.0000921) | (0.000320) |
| N | 7673 | 7673 | 7673 | 6557 |
| Rsquared | 0.000103 | 0.0000878 | 0.0000893 | 4.12e-08 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the city level
The base category is actual buffer size calculted using regional growth

**4.A.3.2 Cross-Sectional Results** Cross-sectional results for fluvial, pluvial and coastal flood risks (these tables refer to the IV presented in the main analysis using the auto-regressive regional growth rate).

Table 4.A.6: New share SE in HFR in cross-sectional OLS (Fluvial flood risk)

| | (1) | (2) |
|---|---|---|
| | 1985 | 2000 |
| Share HFR | 0.988** | 1.053** |
| | (0.0586) | (0.0538) |
| Share unviable (steep) | 0.183** | 0.177** |
| | (0.0334) | (0.0255) |
| Share unviable (rivers) | -0.0225 | 0.0807 |
| | (0.0390) | (0.0628) |
| N | 1771 | 1771 |
| Rsquared | 0.710 | 0.696 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Table 4.A.7: New share SE in HFR in cross-sectional OLS (Pluvial flood risk)

| | (1) | (2) |
|---|---|---|
| | 1985 | 2000 |
| Share HFR | 1.120** | 1.345** |
| | (0.176) | (0.223) |
| Share unviable (steep) | 0.0987** | 0.0838** |
| | (0.0184) | (0.0223) |
| Share unviable (rivers) | -0.0186 | -0.0138 |
| | (0.0101) | (0.00905) |
| N | 1732 | 1732 |
| Rsquared | 0.517 | 0.487 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

Table 4.A.8: New share SE in HFR in cross-sectional OLS (Coastal flood risk)

| | (1) | (2) |
|---|---|---|
| | 1985 | 2000 |
| Share HFR | 0.307 | 0.471* |
| | (0.151) | (0.180) |
| Share unviable (steep) | 0.384* | 0.221* |
| | (0.136) | (0.0939) |
| Share unviable (ocean) | 0.186 | 0.193 |
| | (0.0897) | (0.0904) |
| N | 190 | 190 |
| Rsquared | 0.252 | 0.422 |

* $p < .05$, ** $p < .01$
Standard errors clustered at the region level

**4.A.3.3 First Stage Results** First-stage results for fluvial, pluvial and coastal flood risks (these tables refer to the IV presented in the main analysis using the auto-regressive regional growth rate).

Table 4.A.9: First stage (Fluvial flood risk)

| | (1) Δ Share Share HFR (fluvial) | (2) Δ Share unviable (steep) | (3) Δ Share unviable (rivers) |
|---|---|---|---|
| Δ Share HFR | 0.311** | -0.00846 | 0.0258** |
| IV (fluvial risk) | (0.0383) | (0.0151) | (0.00926) |
| Δ Share unviable | 0.00341 | 0.293** | 0.00865 |
| IV (steep) | (0.0157) | (0.0338) | (0.00908) |
| Δ Share unviable | -0.0147 | -0.0414* | 0.167** |
| IV (rivers) | (0.0182) | (0.0154) | (0.0240) |
| N | 1771 | 1771 | 1771 |
| F | 27.88 | 25.97 | 24.78 |

*$p < .05$, **$p < .01$
Standard errors clustered at the region level

Table 4.A.10: First stage (Pluvial flood risk)

| | (1) Δ Share Share HFR (pluvial) | (2) Δ Share unviable (steep) | (3) Δ Share unviable (rivers) |
|---|---|---|---|
| Δ Share HFR | 0.235** | -0.227* | 0.0505 |
| IV (pluvial risk) | (0.0410) | (0.0906) | (0.0379) |
| Δ Share unviable | 0.000566 | 0.284** | 0.00526 |
| IV (steep) | (0.00392) | (0.0334) | (0.00848) |
| Δ Share unviable | -0.000907 | -0.0417* | 0.177** |
| IV (rivers) | (0.00419) | (0.0157) | (0.0196) |
| N | 1732 | 1732 | 1732 |
| F | 12.45 | 25.52 | 30.92 |

*$p < .05$, **$p < .01$
Standard errors clustered at the region level

Table 4.A.11: First stage (Coastal flood risk)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Δ Share | Δ Share | Δ Share |
| | Share HFR (coastal) | unviable (steep) | unviable (ocean) |
| Δ Share HFR in | 0.133 | 0.00739 | 0.0244 |
| IV (coastal risk) | (0.0681) | (0.0218) | (0.0339) |
| Δ Share unviable | 0.00736 | 0.0716 | -0.124 |
| IV (steep) | (0.0844) | (0.0575) | (0.103) |
| Δ Share unviable | 0.0510 | -0.0139 | 0.256 |
| IV (ocean) | (0.0584) | (0.0228) | (0.125) |
| N | 190 | 190 | 190 |
| F | 3.859 | 4.284 | 3.850 |

$^{*}p < .05$, $^{**}p < .01$
Standard errors clustered at the region level

# 5 Floods and Urbanization in Sub-Saharan Africa

## 5.1 Introduction

The consequences of floods on the loss of human life and destruction of infrastructure are well-documented, but the long-term impacts are less known. According to the Centre for Research on the Epidemiology of Disasters (CRED), most reported natural catastrophes are hydrological disasters (floods, landslides, and tsunamis) followed by meteorological disasters (storm, extreme temperature) as shown in Figure 5.1.1. Floods are responsible for a quarter of economic loss inflicted by natural hazards (Figure 5.1.2). Sub-Saharan African countries are particularly impacted by their geographic locations and their lack of protective infrastructure (Stéphane Hallegatte et al. 2020b; IPCC 2012; Samson et al. 2011). Reconstruction funds are limited in space and time. That is why floods' human cost and share costs per GDP are disproportionately falling on low-income countries (Figure 5.1.3). From 1985 to 2019, floods worldwide killed more than 650,000 people and displaced over 650,000,000 people (R. Brakenridge 2019).

These figures do not take into account the aftermath of floods. In the short-term, they are often followed by disease outbreaks such as malaria, diminishing resilience ability of the poorer households and displacement of populations (McMichael, Barnett, and McMichael 2012; Elsanousi et al. 2018; Boyce et al. 2016; Jones, Ballon, and Engelhart 2018; Ghimire, Ferreira, and Dorfman 2015).

In the long term, it has been found that households hit by floods have poorer health outcomes, lower achievement at school for children and a lack of ability to invest (Lomnitz 2015). Floods have a large impact on agricultural production and livelihood (Dell, Jones, and Olken 2014; Hochrainer 2009). Nonetheless, there is a growing number of people living in high flood-risk areas (Stéphane Hallegatte et al. 2020a). Large floods do not result in a long-term population adjustment in urban areas (Kocornik-Mina et al. 2018): people come back to flooded areas in spite of the apparent dangers.

There is also some evidence that people locate in riskier areas (Rappaport and Sachs 2003;

G. a. Kahn and Benolkin 2007; Pielke et al. 2008). It has been hypothesized that it might be either because these areas are more productive or because the local population does not bear the full costs of floods' destruction. Reconstruction and protection efforts in impacted areas in the US have been pointed out to explain these surprising results (Boustan et al. 2012, 2020). The context of Sub-Saharan Africa could be different as local governments have a more limited ability to reconstruct and support impacted populations and businesses.

In this paper, I investigate if there is any persistent effect of large floods on population growth in Sub-Saharan Africa. The main contribution of this paper is to show that the long-term impact of floods on population growth is different between urban and rural areas. To this end, I observe floods and population growth for three periods (1975-1990, 1991-2000, and 2000-2015), which allow quantifying persistence for a 10-15 years window. Population growth matters as it has been linked to the aggregate economic growth rate. In particular, in the case of Sub-Saharan Africa, the high ratio of workers relative to dependents create opportunities for economic growth. The impact of floods on population growth rates in Sub-Saharan Africa opens the discussion on managing flood risks. As in richer countries, floods could lead to an increase in population growth due to disaster relief in flood-prone areas. Another possibility is that once destroyed infrastructures have been rebuilt, cities would go back to their "natural" growth rate, ignoring the signal to stay away from dangerous areas. Finally, because of the repetitive destruction of means of production or population updating their risk perception, population growth rates could suffer from a long-lasting decrease with the richest and more resilient households moving away.

To understand the link between floods and long-term population growth, I use a simple model where the incidence of floods affects households' decisions to migrate. An increase in the number of flooded days that repetitively destroy means of production in the medium term gives people an incentive to move away or avoid an area. It might lower agglomeration economies enough to impact affected areas persistently.

One of the main problems of studying floods in Sub-Saharan Africa is the lack of national flood risk maps and detailed histories of past floods. Until recently, the studies of the impacts

of floods have been concentrated on case studies using detailed maps of flooding, generally only available for high-income countries or major cities. The second approach consists in using global datasets such as the Emergency Events Database (EM-DAT) or Dartmouth Flood Observatory (DFO). These datasets are constructed using news articles, official reports, and self-reporting. Both of these datasets' coverage and measurement quality depend on the economic and political conditions of where the events happen (Hsiang and Jina 2014). For these datasets, there is a clear improvement in reporting through the period studied, meaning that floods happening in earlier years or more remote areas are likely to be under-reported. To solve this problem, I use river discharge analysis for large river basins created using historical climate simulation and a flood model for the world from 1979 to today (GloFAS) (Alfieri et al. 2013). This data is used to fill gaps in the historical recording. It has the advantage of being a comprehensive coverage of floods for the period, not biased by economic and political conditions. Even though the floods obtained are historical predictions from the model, I refer to them as floods in the rest of the paper for simplicity.

My main variable of interest is long-term population growth in Sub-Saharan Africa. I use the Global Human Settlement Layer (GHSL) dataset to construct a panel data for each 0.1 by 0.1-degree grid cell (about 11km at the equator) for three periods: 1975 to 1990, 1991 to 2000, 2001 to 2015. The GHSL dataset is the most disaggregated population panel available in the world. It also offers a built-up dataset and classifies cells by degree of urbanization. The built-up provides a measure of human activity at the grid cell level. The degree of urbanization is crucial to the analysis as it divides the space into uninhabited, rural, small cities and large agglomerations. It is provided using built-up data from LANDSAT for the four epochs (1975, 1990, 2000, and 2015) and allocating population from census data to each cell. The cut-off used to distinguish between urban and rural areas is 300 inhabitants per km square and a minimum of 5,000 inhabitants. These two criteria allow comparison through time.

The data show that flood risk varies considerably over time. The population growth rate decreases by five percentage points for each additional day of flooding per year. The effects linger even after 10 to 15 years with a growth rate decrease of 10 percentage points. However, in urban areas, the effects are partially dissipated. In large cities, floods have almost no

long-term effect on population growth.

To obtain a more detailed estimate of the temporal effect of floods in urban areas, I analyze the lag effects of past floods on night lights. I find that the impacts of large floods are perceivable up to 5 years after a flood.

Finally, I perform supplemental analysis on Demographic and Health survey data to document the channel between rural and urban differences in floods' impact. Rural households are more likely to be in the lowest wealth quantile after large floods. They also have worst health outcomes. However, most of the differences are driven by the sorting of the poorest people in risky areas.

This paper contributes to the literature on the long-term impact of natural catastrophes in Sub-Saharan Africa, as well as the persistence of cities after large shocks. There is evidence that floods contribute to lower population growth and increased poverty rates in rural areas, primarily due to sorting. However, in urban areas, there is little persistence of floods.

The rest of the paper is structured as follows. First, I review the literature (Section 5.2), then I summarise the context and data (Section 5.3), and the conceptual framework (Section 5.4). I present next the empirical strategy and results (Section 5.5), and in the last part, I conclude (Section 5.6).

## 5.2 Literature review

The two key strands of literature relevant to this paper are (1) the economic consequences of natural disasters and climatic shocks and (2) urban growth dynamics in developing countries.

First, this paper contributes to the literature on the impact of natural disasters as reviewed by Cavallo and Noy (2012). While short-term impacts are well-known, the literature on long-term impacts is relatively recent. There is growing evidence that natural disasters can have a long-term effect on macroeconomic growth. Hsiang and Jina (2014) finds evidence that cyclones lower national incomes relative to their pre-disaster trend and do not recover within

Figure 5.1.1: Natural catastrophes 1980-2019

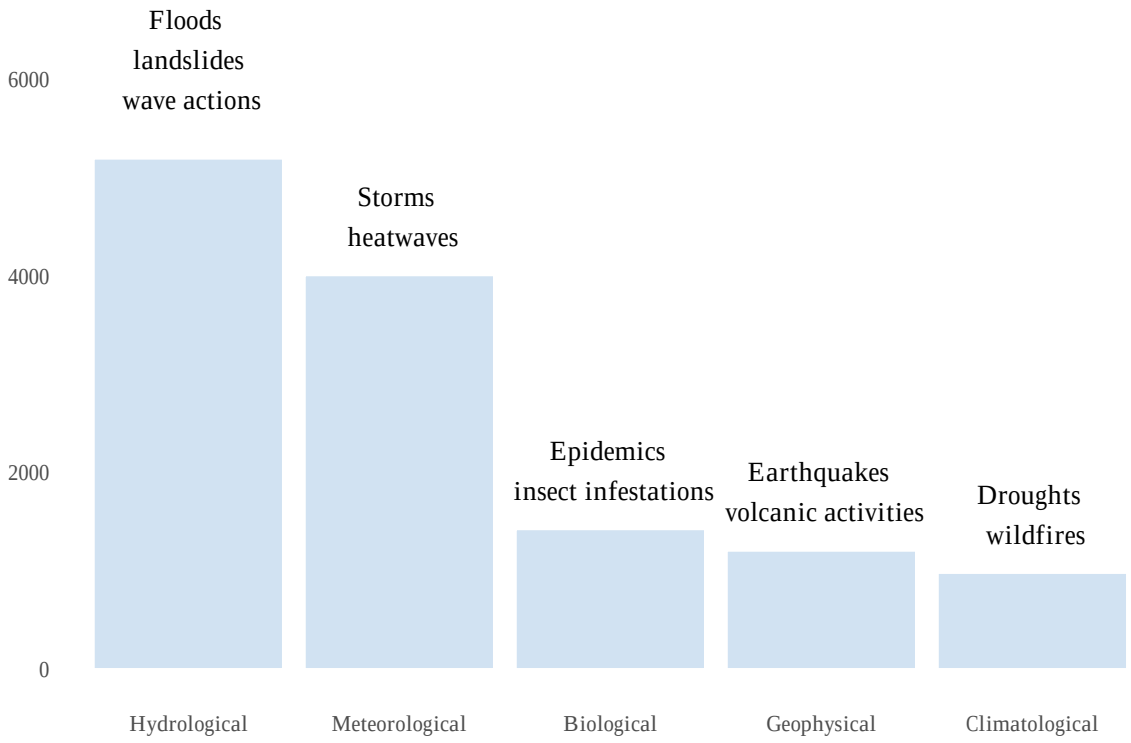
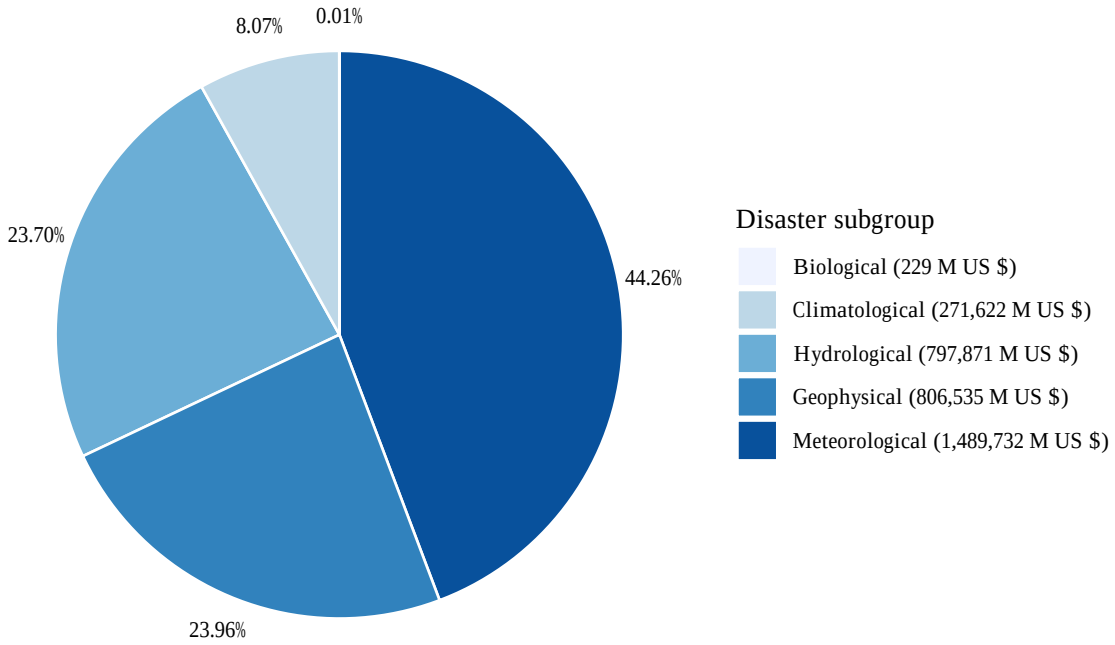
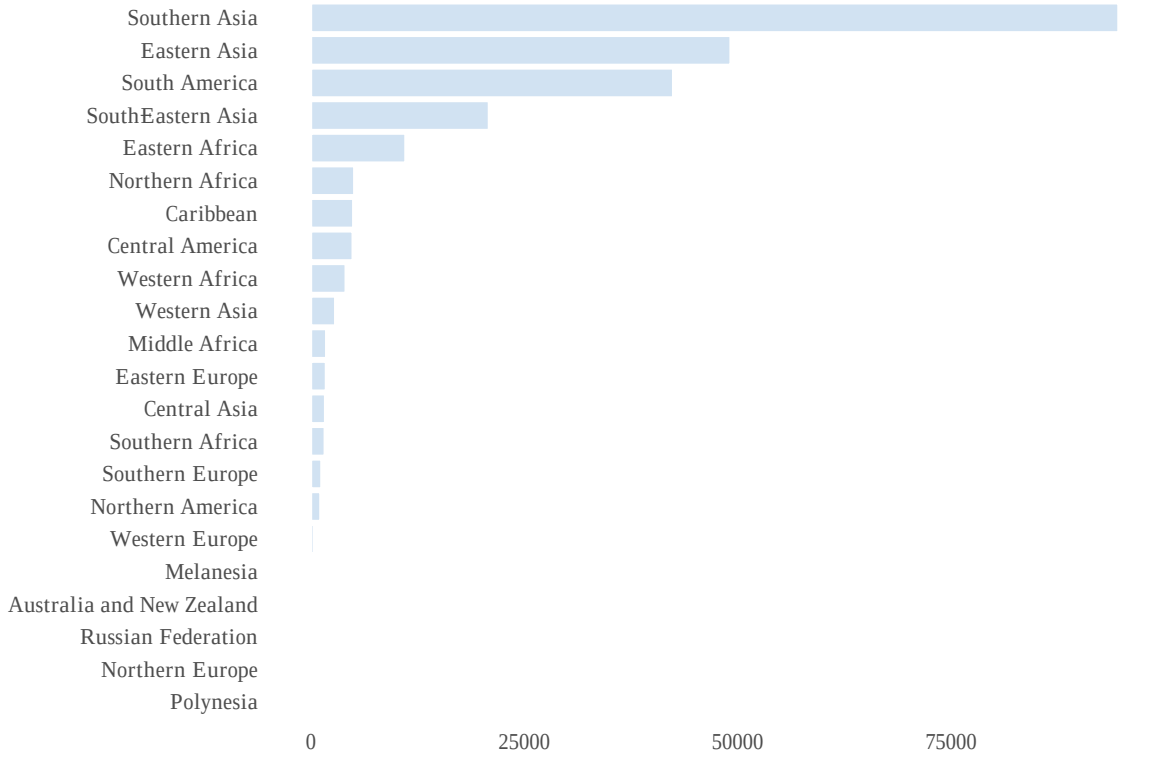



Figure 5.1.2: Natural catastrophies total damages 1980-2019




Source: EM DAT, natural disasters from 1980 to 2019

Figure 5.1.3: Natural catastrophies total death 1980-2019




Source: EM DAT, natural disasters from 1980 to 2019

twenty years.

The level of development of countries influences the time-persistence: McDermott (2012) finds that in countries with low levels of financial sector development, natural disaster events negatively impact economic growth. McDermott, Barryy, and Tol (2014) uses the reported floods from the international disasters database (EM-DAT) to show that in low-income countries characterized by an absence of financial sectors, natural disasters have persistent adverse effects on economic growth over the medium term. Castells-Quintana, del Pilar Lopez-Uribe, and McDermott (2015) also find that poorer countries and households are more vulnerable to climate change-related events.

In addition, linking natural disasters to urbanization, floods contribute to migration to cities, which could fuel ethnic violence and conflicts (Castells-Quintana and Mcdermott 2019). It also suggests that people living in informal settlements are more vulnerable to adverse effects than the rest of the population. Narloch and Bangalore (2018) find that in Vietnam, poverty

levels are higher in regions more exposed to natural disasters, and the poorest households are the most impacted. In addition, they find differences between urban and rural areas in terms of consumption rate. I find a similar pattern in Sub-Saharan Africa.

Institutions have been cited as a major factor entering the consequences of natural disasters. In a country with poorer institutions, citizens are likely to mistrust government efforts; governments are also less likely to seek to improve the welfare of their citizens (M. E. Kahn 2005; Toya and Skidmore 2007; Raschky 2008). On floods, in particular, Besley and Burgess (2002) finds that in India more accountable politicians (via newspapers) are more likely to prevent and mitigate the impacts of a shock.

In the US, government subsidies and private insurance are likely to make people more prone to move to more productive but riskier areas. Deryugina (2011) finds that public and private transfers absorb most of the costs of hurricanes in the US. However, she finds no change in the population, earnings, or employment. She concludes that there are large inefficiencies due to taxation deadweight loss and moral hazard. Boustan et al. (2012) find that young men move away from areas hit by tornadoes but are attracted to areas experiencing floods in the US in the 20s and 30s, potentially due to efforts by the Army Corps of Engineers to protect against future flooding. Governmental investments in protective infrastructure and insurance might lead people to not invest in self-protection (Peltzman 1975; Kousky et al. 2006). To explain these reversed results, Kellenberg and Mobarak (2008) shows a non-linear relationship between development and environmental risks: a higher level of economic development is associated with riskier behaviour (locations near coasts and floodplains). Looking at more long-term outcomes, Boustan et al. (2020) find that severe natural disasters increase out-migration and falling house prices in the US in the twentieth century. They also observe an increase in poverty rates in affected counties; they hypothesize it might be due to the out-migration of richer households, the in-migration of poorer households due to falling out-prices or the pauperization of existing households. Deryugina (2017) also find an increase in unemployment and medical insurance after a disaster.

In general, evidence of the impacts of floods on developing countries is more scarce. In

their studies of large floods, Kocornik-Mina et al. (2018) do not find any persistence of floods on night light activity, except in newly lit areas. Scussolini et al. (2016) finds that people in newly urbanized areas are particularly vulnerable to flood risk.

On migration specifically, Bohra-Mishra, Oppenheimer, and Hsiang (2014) shows that climate migration is more likely triggered by long-term change such as an increase in temperature or precipitations than a one-off catastrophic event. One main reason is that disasters can reduce households' ability to send migrants or relocate by destroying assets.

Finally, these papers relate to the importance of agglomeration economies in urban persistence. Bleakley and Lin (2012) find that cities develop near-natural features that give them a productivity advantage. Cities persist even after these advantages disappear, showing strong path dependency due to agglomeration economies. Unfortunately, it also means that unaccounted future changes in risk can lead to sub-optimal allocation of firms and people across space. Balboni (2021) shows that road investments in Indonesia do not consider climate change and lead to such misallocation. Different forces drive human settlements in high flood risk areas. First, original settlements are often close to flood risks as flood plains, rivers and coasts gave a natural advantage to these locations. Then, agglomeration economics can lead to a higher growth rate in these areas, even if the original advantage has disappeared. Finally, cities can continue to grow along flood-risk areas, for example, along coastlines, as the benefits outweigh the risks. One issue in looking at population growth and flood risk is thus disentangling the unobserved amenities and productivity factors from the floods' direct impact.

Aside from natural disasters, large events such as bombing and war destruction have been studied intensively in the US, Germany, and Japan without finding any long-term impact on the systems of cities (Davis and Weinstein 2002). Mass destruction of parts of a city has even been found to foster growth and reconstruction incentives (Richard Hornbeck et al. 2017). However, the context of Sub-Saharan Africa is hugely different: the original system of cities was constructed under colonialism with the main goal of extracting resources to bring them to Europe (Jedwab et al. 2014). It is thus less likely that the original forces that created the

system of cities are still in place. A major disaster could lead cities to a different growth path. Moreover, agglomeration economies in the context of rural production or informal activities in cities are not well-understood. It is thus unsure that major disruptive shocks would have the same impact as in previous studies of the same kind.

## 5.3 Data and context

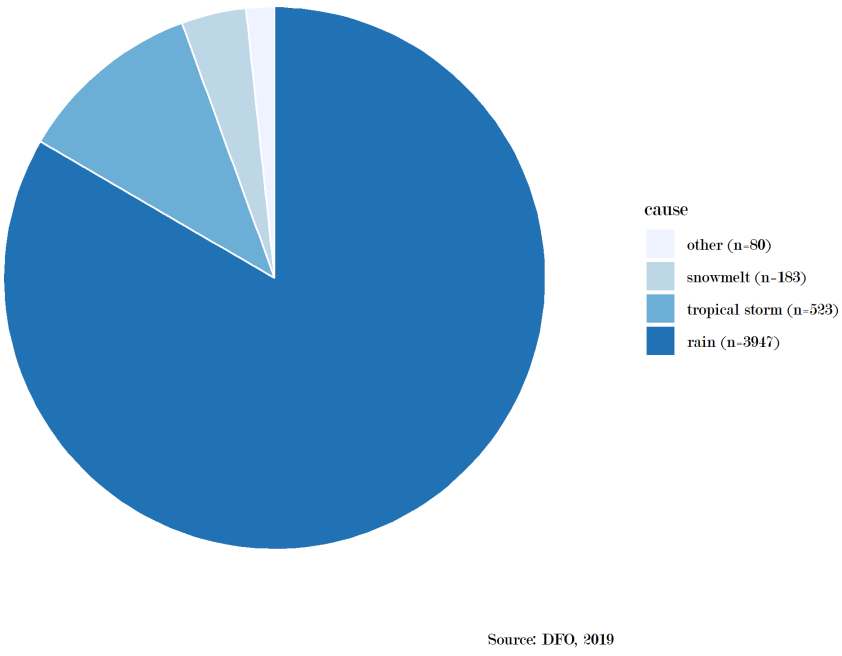
### 5.3.1 Flood data sources

This paper aims to analyze the long-term impacts of large floods on population growth rates. To this aim, I used predicted historical floods from the Global Flood Awareness System (GloFAS), a global hydrological forecast and monitoring system. To capture long-term trends in population growth, I use the Global Human Settlement Layer (GHSL), a gridded population dataset based on satellite imagery and census data. I also use Night Light satellite imagery to explore the temporal dimension of flood persistence at a yearly frequency. Finally, I add to the analysis by measuring large floods' impact on households' wealth and health using Demographic and Health surveys for Sub-Saharan Africa.

Floods can be caused by different events as shown in Figure 5.3.1. This paper focuses on infrequent fluvial flooding in large river basins. I use predicted historical data - so my measure of floods come from historical meteorological conditions and geographic setting. They do not capture changes in land use or political and economic conditions. Compared to reported floods, I am unlikely to bias my results by capturing dominant city effects or anthropogenic changes. However, it also means that I am excluding other flood risks such as surface flooding which are quite common in urban areas and mostly due to land-use changes (Milly et al. 2002; Scussolini et al. 2016; Kundzewicz et al. 2014; IPCC 2012). I also won't consider flood defences and improvement in flood risk management. It could lead to under-estimate the impact in most protected areas. Therefore, the results of the population change's regressions presented below could be an under-estimation of the real impact of flooding.

In the analysis, I use a global hydrological model to predict floods in large basins (GloFAS) jointly developed by the European Commission and the European Centre for Medium-Range

Figure 5.3.1: Main causes of flooding




Source: DFO, 2019

Weather Forecasts (ECMWF).[27] The main goal of Glofas is to issue warnings to support preparatory measures for flood events worldwide. But the model is also used to develop river discharge predictions for the past 40 years using reanalysis climate data. [28]

As the model uses reanalysis data, I can compare these daily river discharge reanalysis time series to actual discharge measures from gauge data[29]. The model has been calibrated to fit the data and reflects the real gauge measures. In Figure 5.3.2, I plot the predicted daily discharges from the GloFAS model to the reported daily discharges in $m^3/s$ from the Global Runoff Data Centre (GRDB), a global hydrological data set that comprises discharge data of more than 9,900 gauging stations from all over the world.

I have 241 stations in Sub-Saharan Africa over 13,439 days (the panel is unbalanced). The correlation coefficient between the reported discharges from GRDB and the predicted discharges from GloFAS is 0.7835, and the p-value is $2.2e^{-16}$. The model seems to be able to predict between 30 and 55% of the variation in predicted versus reported discharges between the different periods. On average, a 10% increase in reported river discharge is associated with an 8.8% increase in predicted river discharge. The model underpredicts discharge slightly compared to reported gauge data, but there is still a strong correlation between the two datasets.

To get floods from river discharges, I compare the daily discharge to discharge thresholds for each cell also provided by GloFAS[30]. Discharge thresholds are a set of maps corresponding to different floods' return periods (2, 5, and 20 -years, respectively called Medium, High and Severe Awareness Levels). The return period is a measure of the probability that a flood will occur. For example, the return period of a flood could be 20 years, meaning that the probability is 1/20, or 5% in any one year. The thresholds have been created using the long

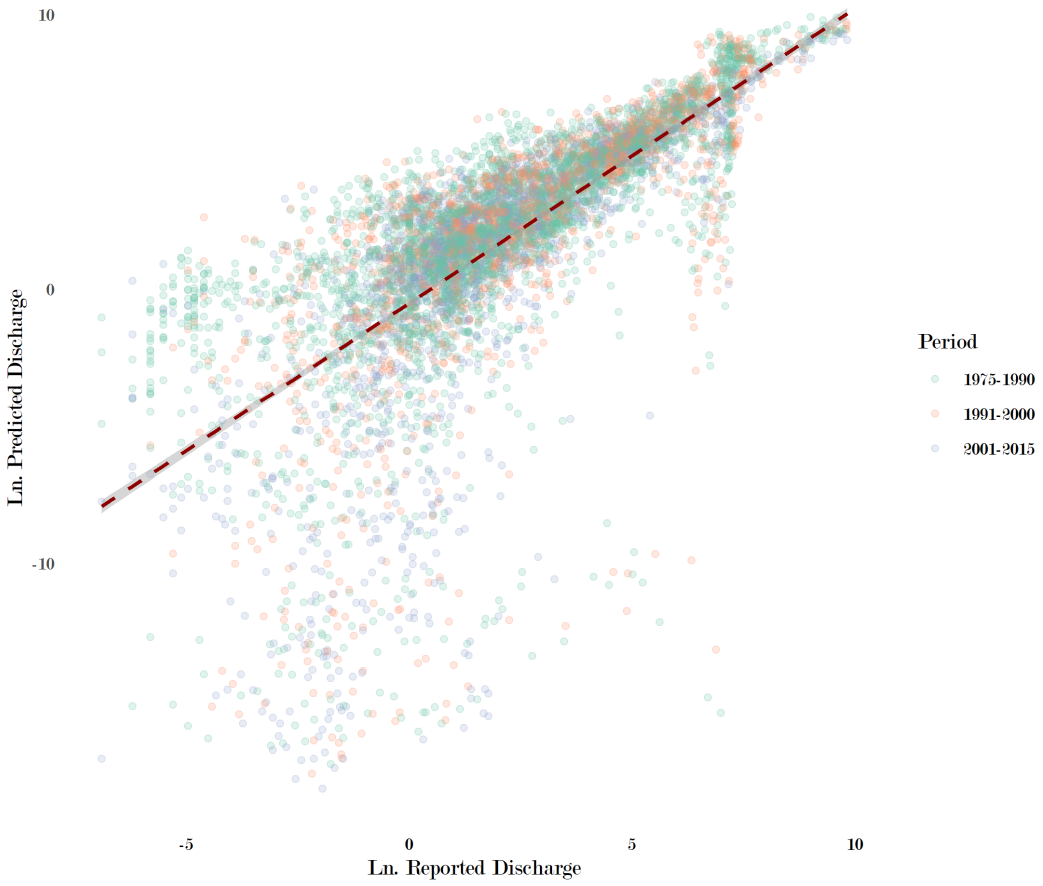
[27] GloFAS uses daily meteorological forecasts and hydro-meteorological initial conditions to feed the hydrological modelling based on a land surface model (HTESSEL) and a flood model routine (Lisflood).

[28] The GloFAS v2.0 river discharge reanalysis was produced using the ERA5 ECMWF reanalysis (ERA5pr). Reanalysis data is used to predict most accurately past weather and climate conditions. It is created using past weather observations with modern forecasting models. One of the strengths is to provide consistent and global coverage.

[29] A stream or river gauge station is a permanent point where the water level or flow of water is measured.

[30] "GloFAS 30-day discharge thresholds are a set of maps of discharge magnitudes corresponding to the 2-, 5- and 20-year return period floods (respectively called Medium, High and Severe Awareness Levels). They are generated by fitting a Gumbel extreme value distribution to the annual maxima series extracted from GloFAS 30-day discharge time-series of the hydrological reanalysis using the method of L-moments."

Figure 5.3.2: Reported versus predicted river dicharge



reanalysis time series. The GloFAS dataset uses climate data from 1979 to today, so it itself spans 40 years. The system is presented in the Appendix Figure 5.A.1.

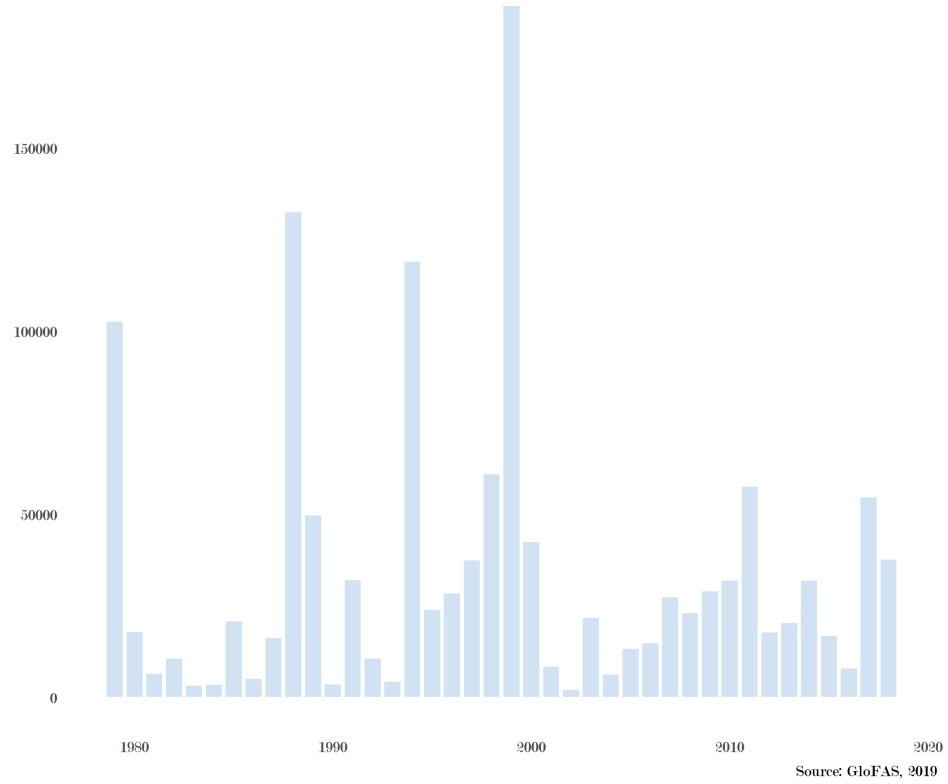
One of the main advantages of GloFAS over other flood sources is that it is independent of administrative and political boundaries and is consistent through time. The dataset covers large basins around the world from 1979 to today. While it does not rely on reported events, any bias introduced by the model is unlikely to be linked with the short-term economic conditions of where the floods happen. On the other hand, one of the main limitations of GloFAS is that the model is only operating on large river basins (upstream areas $>$ 10,000 km square) and only reports on rain-fed fluvial floods. Floods caused by other factors such as coastal erosions are not taken into account. The model also performs worst in arid and semi-arid regions, particularly in Australia, Mexico and in the Sahel, but large floods are less impacted (Alfieri et al. 2013).

In Figure 5.3.3, I aggregate for all cells in Sub-Saharan Africa the number of times their discharge exceeds the 20-year return period threshold. Contrary to other datasets collected through newspapers and official reports, there is no clear time trend. There is, however, quite a bit of variation - with about 150,000 cells by day flooded in 1999 and only 1,400 three years later in 2002. In addition, the last period has fewer predicted flooded days than previous periods, which is consistent with a decrease in moisture in Africa. The variability and overall decrease seem to indicate that the results are not driven by a medium-term change in risk perceptions. In the appendix, I present a comparison of GloFAS floods measure to the Dartmouth Flood Observatory.

### 5.3.2 Population data sources

The population data comes from the Global Human Settlement Layer dataset. It uses satellite imagery, census data, and GIS data to create a dataset on population (GHS POP), urbanization (GHS-SMOD) and built-up land (GHS BUILT-UP) for the whole world. It is available for four periods (1975, 1990, 2000 and 2015) at a 1x1 km uniform global grid resolution. Compared to national datasets on population and urban growth, the GHSL methodology has a consistent measure for different countries. Variation is only introduced by different years of capture in

Figure 5.3.3: Floods in Sub-Saharan Africa




the satellite imagery[31] or censuses[32]. It is also more disaggregated than other population datasets. The system is schematized in the Appendix Figure 5.A.6. However, data quality has increased over time as satellite imagery has gotten more precise and censuses more frequent. To mitigate this issue, I dropped the first period in the robustness check and used different thresholds to define recently populated areas.

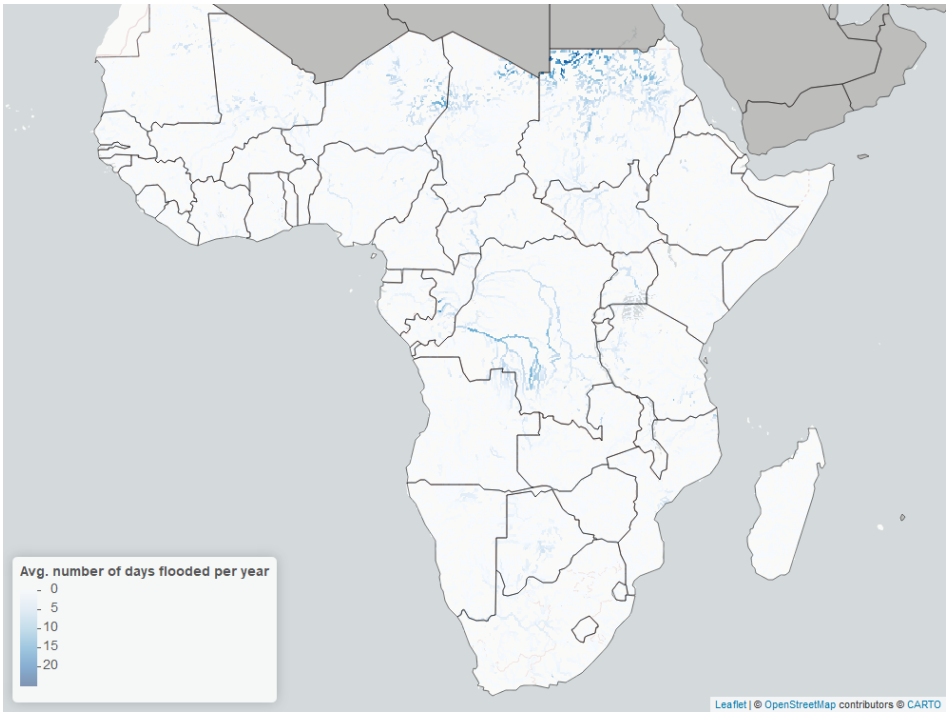
One of the advantages of the GHSL dataset is that urban thresholds are consistent over countries and time. The urbanization grid has four different categories: uninhabited, rural, towns and suburbs or small urban areas, and cities or large urban areas. To be classified as towns and suburbs or small urban areas, the cell needs at least 300 inhabitants per km2 and a cluster of a minimum of 5,000 inhabitants. Cities or large urban areas are a set of contiguous cells with a density of at least 1,500 inhabitants per km2 or a density of built-up greater than 50% and a minimum of 50,000 inhabitants.

[31]The built-up data comes from the Landsat multitemporal collection
[32]The population dataset comes from CIESIN GPWv4

One of the main challenges of analyzing the impact of floods on population growth is to compare floods - temporarily and spatially concentrated in a small space and time, to population data which is usually available every ten years.[33]

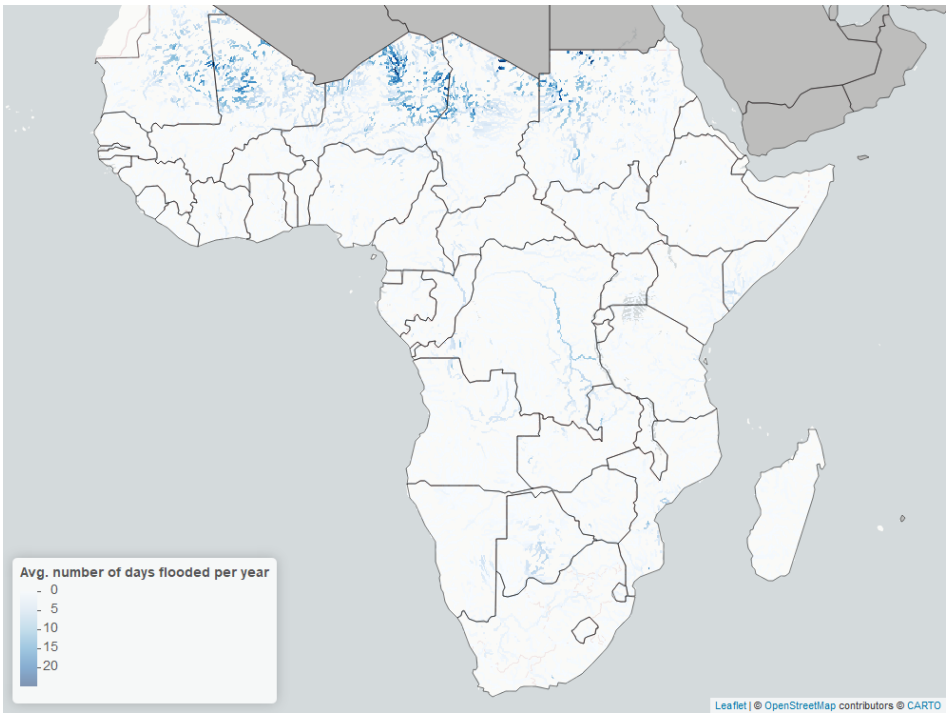
I use the daily discharge grids and the 20-year return threshold grid to construct my main measure of floods. I consider a cell flooded when the daily discharge exceeds the 20-year threshold (which is the largest return threshold consistently available). To obtain a measure that I can link with population growth, I sum up each day flooded in a year for each cell to get a yearly sum. I then average each year in a period to get the average yearly number of flooded days per period (the three periods are 1979 to 1990, 1991 to 2000, 2001 to 2015). The results can be shown in maps 5.3.4, 5.3.5, and 5.3.6. Sub-Saharan Africa is shown in white background. While being a simple measure of the frequency of floods in the area, it smooths out some interesting variations about how large or how long each flood is.

Figure 5.3.4: Average number of flooded days per year, SSA, 1979-1990




[33]To compare the floods and the population datasets, I aggregate the population data recorded to the 1 by 1 km grid cell to 0.1 by 0.1 degrees. It requires summing up the data by a factor of about ten and then transforming the spatial references. For the urban settlement, I first recode the urban settlement variables to null if the cell is uninhabited or rural, and one if it is a town/suburb or a city/urban large urban area. I then create an average share of urban cells when I aggregate up. The process introduces some measurement errors in the population and urban data.

Figure 5.3.5: Average number of flooded days per year, SSA, 1991-2000




### 5.3.3 Data description

In the sample I use for the analysis, I obtain about 125,000 cells for each of the 44 countries in Sub-Saharan Africa. I only include cells that are inhabited for the three periods. I present the summary statistics for all cells in SSA in Table 5.3.2, for cells with any urban population in 1975 in Table 5.3.3 and for cells without any urban population in 1975 in Table 5.3.4. I find that rural cells are more exposed to floods than urban cells. It does not change over time. It might be linked to the type of floods used (fluvial risk and not coastal or pluvial), but also it might reflect the patterns of urbanization in SSA.

The average population in each cell is about 5,110 inhabitants. The average annualized population growth rate is 4% from 1975 to 2015. The urban share corresponds to the share of the surface of the cell that is either classified as a town/suburb or a city/large urban area. The average urban area in each cell is about 1% over the three periods and only 0.5% in 1975. However, the urban population share is much larger - 23% over the three periods and 16.4% in 1975.

Figure 5.3.6: Average number of flooded days per year, SSA, 2001-2015


Figure 5.3.7 shows the distribution of the urban population share (for cells with at least some urban population) in 1975. More than three-quarters of the cells do not have any urban population.

The average number of days flooded per year is 0.5. Contrary to the other source of data, the number of floods per year in each cell is stable or decreasing. As the model is based on climate data (most sub-Saharan countries are becoming arider during that time) and other static inputs, it explains the current results. The flood measure only counts cells as flooded on the river line. I construct a 30 km buffer around each cell to capture cells that are flooded around the river and apply a gaussian decay to give a smaller weight to faraway cells. Cells have an average of 1 flooded day per ten years.

### 5.3.4 Demographic and Health Surveys (DHS)

Demographic and Health surveys collect data on women's health, family planning, and child mortality from over 90 countries and 300 surveys. I use a subsample for Sub-Saharan countries with GPS data on clusters' locations. In total, I obtain 138 surveys in 34 countries as shown

Figure 5.3.7: Histogram of share urban population in 1975 > 0

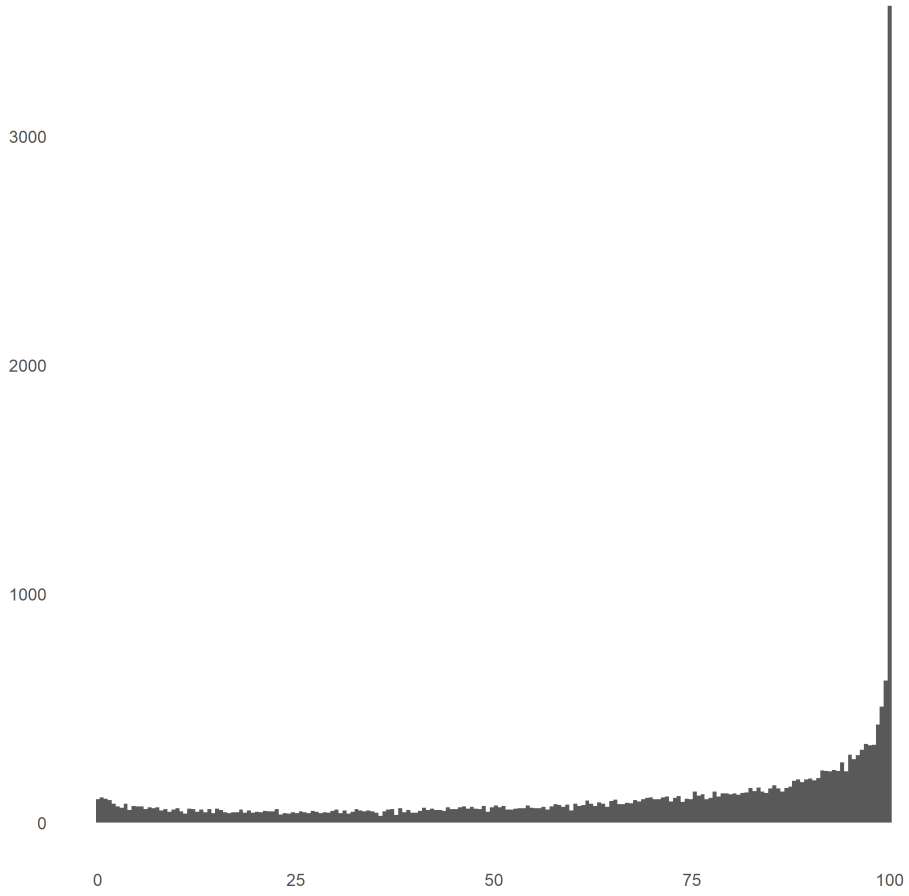



Table 5.3.1: Summary statistics for SSA (time invariant variables)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Population 1975 | 98,184 | 2,978 | 11,535 | 0 | 1,011,443 |
| Urban population share in 1975 | 98,184 | 17 | 34 | 0 | 100 |
| Dist. to the nearest city in 1975 | 91,793 | 201 | 193 | 0 | 1,274 |
| Dist. to the capital city | 98,026 | 486 | 357 | 1 | 1,933 |
| Dist. to the coast | 98,184 | 546 | 410 | 0 | 1,800 |
| Dist. to the nearest borders | 98,184 | 118 | 101 | 0 | 621 |
| Elevation | 98,184 | 719 | 536 | −535 | 4,485 |

Table 5.3.2: Summary statistics for SSA (panel)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Population | 377,602 | 5,114.4 | 21,274.5 | 0.000 | 2,205,667.0 |
| Urban population | 377,602 | 3,910.1 | 20,889.0 | 0 | 2,205,463 |
| Rural population | 377,602 | 1,204.3 | 2,007.7 | 0.0 | 26,509.1 |
| Urban population share | 377,602 | 23.0 | 36.6 | 0 | 100 |
| Population growth (%) | 339,264 | 3.8 | 8.5 | −166.9 | 158.6 |
| Urban population growth (%) | 94,812 | 3.2 | 6.9 | −106.0 | 103.2 |
| Rural population growth (%) | 325,144 | 3.2 | 9.7 | −186.4 | 156.3 |
| Built up share | 377,602 | 0.4 | 2.5 | 0.0 | 103.2 |
| Built-up growth (%) | 257,514 | 7.4 | 8.4 | −58.4 | 136.2 |
| Discharge | 91,696 | 4,969.4 | 29,707.8 | 0.000 | 657,944.2 |
| Flooded days | 91,696 | 0.5 | 1.3 | 0.0 | 21.4 |
| Flooded days buffer | 377,602 | 0.1 | 0.4 | 0.0 | 9.4 |

Table 5.3.3: Summary statistics for SSA (any urban population in 1975)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Population | 70,232 | 19,161.3 | 44,747.1 | 0.001 | 2,205,667.0 |
| Population growth (%) | 70,229 | 2.1 | 4.2 | −112.4 | 77.2 |
| Urban population | 70,232 | 16,566.2 | 44,544.7 | 0.0 | 2,205,463.0 |
| Rural population | 70,232 | 2,595.1 | 2,625.8 | 0.0 | 26,509.1 |
| Urban population share | 70,232 | 66.4 | 32.3 | 0.0 | 100.0 |
| Built up share | 70,232 | 1.5 | 5.4 | 0.0 | 103.2 |
| Built-up growth (%) | 69,803 | 6.5 | 6.8 | −6.1 | 82.6 |
| Discharge | 17,175 | 8,339.1 | 41,105.3 | 0.000 | 657,944.2 |
| Flooded days | 17,175 | 0.3 | 0.9 | 0.0 | 12.2 |
| Flooded days buffer | 70,232 | 0.1 | 0.3 | 0.0 | 5.9 |

Table 5.3.4: Summary statistics for SSA (no urban population in 1975)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Population | 235,333 | 2,229.0 | 7,263.7 | 0.000 | 1,635,575.0 |
| Population growth (%) | 229,568 | 3.8 | 9.0 | −166.9 | 158.6 |
| Urban population | 235,333 | 1,178.9 | 6,815.2 | 0.0 | 1,635,512.0 |
| Rural population | 235,333 | 1,050.1 | 1,849.1 | 0.0 | 25,800.4 |
| Urban population share | 235,333 | 14.4 | 30.5 | 0.0 | 100.0 |
| Built up share | 235,333 | 0.1 | 0.7 | 0.0 | 74.7 |
| Built-up growth (%) | 147,558 | 7.6 | 8.6 | −58.4 | 136.2 |
| Discharge | 57,284 | 4,175.4 | 26,268.8 | 0.000 | 657,888.9 |
| Flooded days | 57,284 | 0.5 | 1.4 | 0.0 | 21.4 |
| Flooded days buffer | 235,333 | 0.1 | 0.4 | 0.0 | 9.4 |

in Figure 5.3.8. The GPS coordinates correspond to the location of sampled clusters in each country. The clusters are selected to be representative of the households of the main regions in each country as well as the urban-rural divide. In each cluster, households are selected to answer population, health, and nutrition questions. I use the module on malaria and wealth. I use clusters' GPS location to merge it with the flood data over time as shown in 5.3.9. A confidential error buffer has been added to preserve the anonymity of the households. It is thus not possible to differentiate between households that were in the flood zone and households that were not or that have moved in since. However, it means that I am capturing both the direct impact of floods as well as the neighbourhood effects in my estimate.

I construct a measure of malaria prevalence and flood occurrence using the person files as shown in Table 5.3.5 and Table 5.3.6. The table shows the raw unweighted counts. About 30% of the sample is urban. Malaria prevalence is much higher in rural areas compared to urban areas (30% against 20%).

In Table 5.3.7 and Table 5.3.8, I use the DHS households files to compare the probability of being in the lowest quantile of wealth index and exposure to floods. I also calculate different measures of flooding: a count of floods for the past ten years and the time since the last flood. Rural households are much more likely to be in the lowest poverty quantile than urban households (30% against 4%).

Table 5.3.5: Summary statistics for malaria prevalence in rural areas (DHS)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Malaria rate | 200,716 | 0.3 | 0.5 | 0.0 | 1.0 |
| Flooded in the last month | 2,322,924 | 0.03 | 0.2 | 0.0 | 1.0 |
| Months since last flood | 2,322,924 | 74.0 | 79.1 | 0.0 | 447.0 |
| # Flooded | 2,332,048 | 1.3 | 3.0 | 0 | 12 |

## 5.4 Conceptual framework

In this section, I use a simple model by Moretti (2011) based on Rosen-Roback spatial equilibrium to shed light on empirical results. This model is generally used to model shocks

Figure 5.3.8: Number of surveys per country

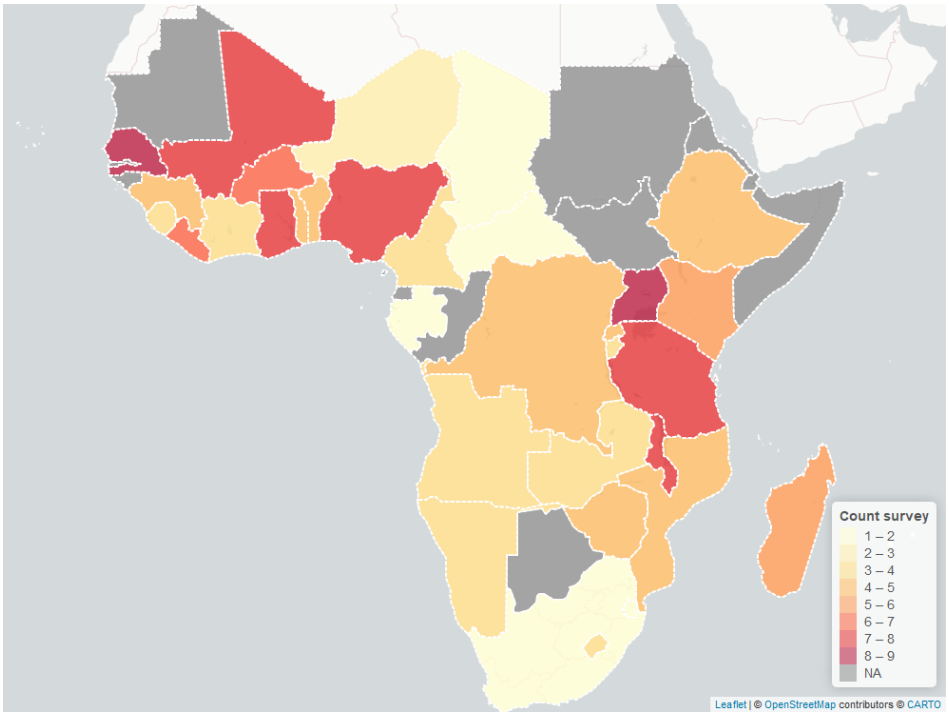



Figure 5.3.9: Number of floods for each cluster

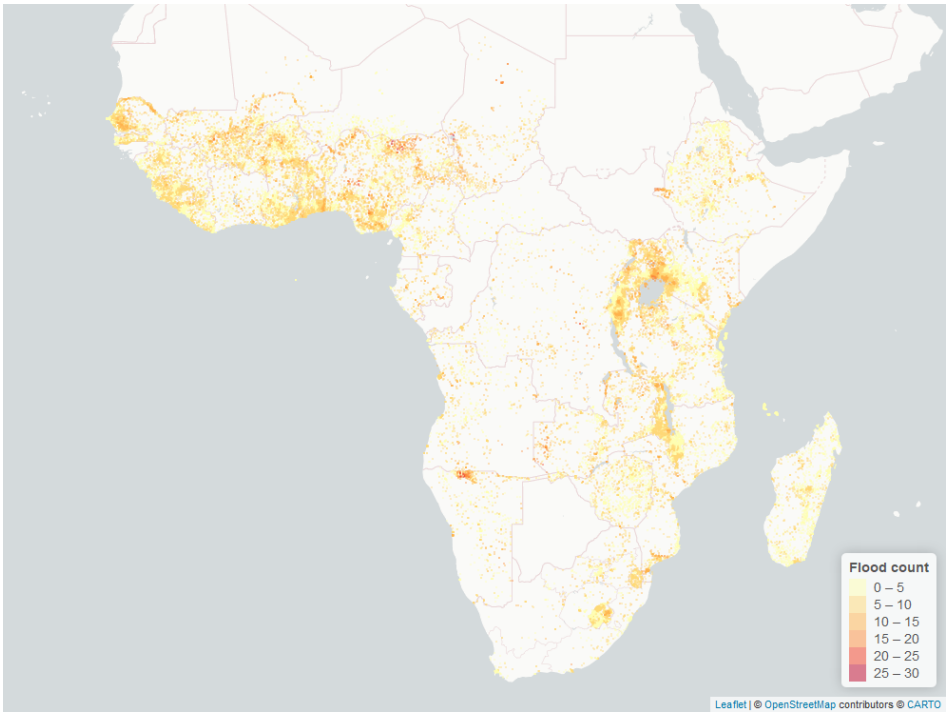


Table 5.3.6: Summary statistics for malaria prevalence in urban areas (DHS)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Malaria rate | 71,943 | 0.2 | 0.4 | 0.0 | 1.0 |
| Flooded in the last month | 1,092,822 | 0.03 | 0.2 | 0.0 | 1.0 |
| Months since last flood | 1,092,822 | 79.6 | 84.7 | 0.0 | 446.0 |
| # Flooded | 1,108,498 | 1.1 | 2.8 | 0 | 12 |

Table 5.3.7: Summary statistics for the wealth index in rural areas (DHS)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Poorest | 715,834 | 0.3 | 0.5 | 0 | 1 |
| Flooded in the last month | 710,221 | 0.03 | 0.2 | 0.0 | 1.0 |
| Months since last flood | 710,221 | 76.7 | 76.5 | 0.0 | 447.0 |
| # Flooded | 715,834 | 1.8 | 1.7 | 0 | 20 |

Table 5.3.8: Summary statistics for the wealth index in urban areas (DHS)

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Poorest | 346,323 | 0.04 | 0.2 | 0 | 1 |
| Flooded in the last month | 339,828 | 0.03 | 0.2 | 0.0 | 1.0 |
| Months since last flood | 339,828 | 83.0 | 83.2 | 0.0 | 446.0 |
| # Flooded | 346,323 | 1.6 | 1.7 | 0 | 20 |

to local economic markets. It is helpful to understand spatial linkage. In the case of floods, the general intuition is that in partial equilibrium, a flood reduces the total population of the affected area. The reduced population put downward pressure on the local housing market, reducing house prices (assuming an inelastic supply). However, the population decrease is partially compensated by people moving in through lower house prices. Richer households are more likely to move out; poorer households might be attracted by lower housing prices.

Using the model developed by Moretti (2011), I assume that all households are similar and provide one unit of labour. They produce a single output which has a unitary price traded everywhere. Workers' locations determine local labour supply. The indirect utility of household $i$ in location $j$ is

$$U_{ij} = w_j + r_j + A_j + e_{ij}$$

Local nominal wage and rent are $w_j$ and $r_j$. Locations $j$ differs by amenities value $A_j$. $e_{ij}$ is an idiosyncratic preference for location $j$.

Now, take two locations $a$ and $b$. The relative preferences for $a$ over $b$ is

$$e_{ia} - e_{ib} \sim U[-s, s]$$

$s$ represents the degree of mobility of the population. It encompasses factors such as attachment to a specific area to mobility costs between different areas. If $s$ is low, labour is mobile.

An household would choose $a$ over $b$ if the locational preferences balance out real wages and amenities differences:

$$e_{ia} - e_{ib} = (w_a - w_b) + (r_a - r_b) + (A_a - A_b)$$

It gives the following labour supply for city $b$ which is controlled by the real wage and amenities as well as the mobility factor $s$. In the case of Sub-Saharan Africa, I expect younger people, and on average a younger population, to have a fairly high labour supply elasticity, thus a

fairly mobile population:

$$w_b = w_a + (r_a - r_b) + (A_a - A_b) + s\frac{N_b - N_a}{N}$$

The production function takes the following Cobb-Douglas form:

$$\ln(y_j) = X_j + hN_j + (1-h)K_j$$

where $X_j$ is a productivity shifter and $K_c$ is the log of capital. The labour demand is the following:

$$w_j = X_j - (1-h)N_j + (1-h)K_j + \ln(h)$$

The housing demand comes from re-arranging the labour supply:

$$r_b = (w_a - w_b) + r_a + (A_a - A_b) + s\frac{N_b - N_a}{N}$$

The supply of housing is a simple function where the housing units equals the number of workers:

$$r_j = z + k_j N_j$$

with $k_j$ the elasticity supply of housing. In this framework, the landowners are assumed to be absentee landlords. I expect the housing supply to be more elastic in rural areas than in urban areas (less regulatory and geographic constraints), and high in general as part of the housing stock is informal. I abstract from ownership and renting discussion at this stage. Ownership is usually higher in rural areas and could also be used to explain why rural areas show a stronger wealth response to floods (destruction of housing would be a major asset loss).

The problem with identifying the impact of floods on the equilibrium is that even with this simple framework, they could potentially enter in three different ways: first, through the amenities of the areas, for example, by destroying public infrastructure; second, through the local productivity shifter by destroying the means of production and finally, through a decrease in the stock of housing.

In the simplest case, there is no flood. At equilibrium, people should sort themselves between locations. I do not expect people to locate in the same place, even if local amenities or productivity are higher, as people have idiosyncratic preferences. Urban areas have higher amenities on average than rural areas. For example, even in slums, access to water, electricity and sanitation is better than in rural areas (Bird, Montebruno, and Regan 2017). Education and health progress have also been faster than in rural areas (Bird, Montebruno, and Regan 2017).

From the simplest case, it is possible to add to the effects of a flood. A flood lowers the population in affected areas through a decrease in amenities and productivity. I choose the simplest entry as a productivity shock in city $b$: $X_{b2} = X_{b1} + \Delta$. I expect the housing elasticity to be fairly elastic as well as labour mobile. Most of the shocks should then be translated by out-migration:

$$N_{b2} - N_{b1} = \frac{N}{N(k_a + k_b) + 2s}\Delta \leq 0$$

The real wages in $b$ are also decreased:

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{k_a N + 2s}{N(k_a + k_b) + 2s}\Delta \leq 0$$

However, this fall in real wage will be partially compensated by a decrease in housing costs:

$$r_{b2} - r_{b1} = \frac{k_b N}{N(k_a + k_b) + 2s}\Delta \leq 0$$

Real wage in $a$, the non-affected region, might also be decreased, but by less than in the impacted area $b$ :

$$(w_{a2} - w_{a1}) = (r_{a2} - r_{a1})\frac{k_a N}{N(k_a + k_b) + 2s}\Delta \leq 0$$

The two key parameters are the housing elasticity $k_j$ and labour mobility $s$. Holding housing elasticity constant, the more mobile the population is, the more they can relocate to reduce the impact of the shock. If $s$ is small, the elasticity of labour supply is high; people are

more likely to relocate after a shock.

Housing supply is directly linked to geographic constraints and the regulatory environment (Saiz (2010)). I expect a more elastic supply in rural areas, so a more considerable population drop and a smaller decline in price/rent. Recent papers by Baum-Snow and Han (2022) and Hilber, Lyytikäinen, and Vermeulen (2011) show indeed that housing supply elasticity depends on distance to city centre and geographic constraints. One caveat of this model is that housing supply is generally represented as kinked downwards. One argument to use this model is that the housing stock is less durable in developing countries, especially in rural areas, so destruction might be more feasible.

Floods could also be modelled as a partial destruction of housing stock instead of being an amenity or productivity shock. It would lead to an increase of increase house prices and rents in both urban and rural areas. However, the destruction could be larger in rural areas (less flood protection, and less sturdy buildings).

I relate this simplest model with my empirical strategy by looking at population changes after a flood. Similarly to Kocornik-Mina et al. (2018), I do not find significant long-term changes in urban areas. It is consistent with urban areas having a more inelastic housing supply.

## 5.5 Empirical analysis and results

In this section, I investigate if there is a lower population growth following a large flood in the medium-term, which would be consistent with out-migration in the affected areas. Next, I look at heterogeneous effects between rural and urban areas as I expect labour and housing elasticities to be different. I then use night lights to measure the response to a severe flood using a yearly measure. Finally, I use demographic and health surveys to understand better how the impacts could differ between rural and urban settings.

### 5.5.1 Population growth

In this section, I present the results from the empirical analysis using the GHSL population data and GloFAS predicted floods.

I am interested in the impact of large floods on population growth in rural and urban areas. To this end, I observe the population in each period (1990, 2000, 2015) and the number of flooded days in the current and previous periods (1975-1990, 1991-2000, 2001-2015). The observation unit is the 0.1 by 0.1-degree grid cell $c$ observed for three periods $t$ for all inhabited cells in Sub-Saharan Africa: 1990, 2000, and 2015. The main identification concern is that cities grow near rivers and takes to benefit from access to clean water sources, agricultural productivity and weather amenities of these places. There could be a reverse causality bias when looking at the impact of floods on urbanization. To alleviate it, I use a fixed-effect model to analyze population changes over time and thus interpret the coefficients on my table as the impact of floods on population change. Precisely, I control for cell fixed effects, country-time trends and period fixed effects in all the population change regressions. However, if flood-prone areas are becoming more and more attractive over time - and there are some indications in the literature that it might be the case in South-East Asia, I might still capture some of the endogenous growth in the regressions. The context in Sub-Saharan Africa is very different to South-East Asia- I do not observe any significant increase in settlements in high-flood risk areas.

Another concern is measurement bias in the dependent variable. The variation I would ideally like to capture is the impact of floods on population growth. However, I observe the predicted number of days flooded for each cell rather than the actual number of days flooded, which themselves are only a proxy for the costs of floods. I assume that the cost of floods is increasing in the number of days flooded in Sub-Saharan Africa. The main independent variable is the average number of days predicted flooded per year divided by 100: $Floods_{c,t}$ in cell $c$ located in the country $cnty$ for the period $t$. To only lead to attenuation bias, the measurement error introduced by the model needs to follow a "classical errors-in-variables model": it needs to have a mean of 0, and not be correlated with the outcome (logged population), the predictors (flood costs, urban share, recently populated and population in

1975) or the error term. As mentioned previously, as the flood model focuses on large flooding driven by climate variables, predicted floods should not be correlated with land-use changes. Contrary to reported floods, it is unlikely to be more precise in areas that have more media attention or where governments are more accountable (e.g. larger cities and political centres).

First, I look at how floods impact rural and urban areas differently. I include the share of the cell classified as urban in the GHSL dataset in 1975 Share urban population$_{c,1975}$. Finally, I control for cell fixed effects, time trends and period fixed effects. All standard errors are clustered at the country level.

The current formulation looks at the relationship between population levels and the average number of flooded days per period. The estimating equation is:

$$\begin{aligned} Ln(Pop_{c,t}) = \alpha_1 \text{Flooded Days}_{c,t} + \alpha_2 \text{Flooded Days}_{c,t-1} + \\ \alpha_3 \text{Share urban population}_{c,1975} \times \text{Flooded Days}_{c,t} + \\ \alpha_4 \text{Share urban population}_{c,1975} \times \text{Flooded Days}_{c,t-1} + \\ + \rho_c + \delta_t + \gamma_{cnty} \times \delta_t + \epsilon_{c,t} \end{aligned}$$

where $Ln(Pop_{c,t})$ is the natural log population in cell $i$ at period $t$. Flooded Days/100$_{c,t}$ is the average number of flooded days per year within 30 km of a cell $i$ at period $t$. Floods in distant cells are weighted less than close cells using a Gaussian decay. The measure of floods is thus similar to an inverse distance weighted average. Share urban population$_{c,1975}$ is the share of the population of the cell $i$ urban in 1975. $\rho_c$ represents cell fixed effects, $\delta_t$ period fixed effects and $\gamma_{cnty} \times \delta_t$ country period fixed effects[34]

In Table 5.5.1, I present the results of the first specification. In column 1, I find a non-significant negative impact of the increase in the average number of flooded days on population growth in Sub-Saharan Africa. In column 2, once I add the number of flooded days in the last period, I find that an additional day of floods in the current periods and the past periods are associated with a decrease in population growth by 3.6% and 5.9%. The population growth

[34]The population is measured in 3 periods of different time lengths: 15, 10 and 15 years. I also run a regression using the difference in the average number of flooded days per period on population growth annualized to deal with the different time length and obtain similar results.

rate over the period is 3.8%, so these areas either do not gain any population or might even lose some. The coefficient is more significant and slightly larger in the lagged term, which is surprising. It might indicate that the coefficients capture some omitted non-desirable aspects of the most exposed areas.

I include the interaction with the share of the urban population in 1975 in column 3. I find that the impact of an increase in flooded days is still negative for rural areas but compensated in urban areas. It seems to indicate that the urban areas do not experience the same loss in population in the medium term.

Once I include the interaction with the share of the urban population in 1975 and lags in column 4, I find a stronger impact on rural areas (a decrease of 5.7 and 7.7 percentage points) but positive interactions with the share urban. It does not mean that urban cells gain population after an increase in floods but that they lose less. Table 5.A.1 in the appendix show exactly that by analyzing the rural sample and urban sample (any cells with an urban population) separately.

Additionally, the more urban a cell is, the more the negative impact of floods on population growth is mitigated. Table 5.A.2, Table 5.A.3, and Table 5.A.4 in the Appendix show that this result is robust to a different definition of urban (taking a dummy for any urban in 1975, a dummy for share urban in 1975 superior to the 90th percentile of the distribution, or a categorical variable taking a value of 0 for rural, 1 for any urban and 2 for share urban in 1975 superior to the 90th percentile). While I do not observe migration directly, it seems to indicate that impacted rural areas lose population and impacted urban areas grow slower than non-impacted cities.

Finally, I perform the same analysis but using the income levels as defined by the UN World Urban Prospect 2018 to create my sample in Table 5.A.5, 5.A.6, 5.A.7, and 5.A.8. I find stronger negative relationships between severe floods and population growth for rural areas in low-income countries (LIC), as well as still a positive interaction with the urban population. This is expected as most of the low-income countries are in Sub-Saharan Africa. I

find almost no statistically significant coefficients for lower-middle-income countries (LMIC) and high-income countries (HIC). I find similar patterns for upper-middle-income countries but smaller coefficients. These findings seem to indicate that the level of economic development and institutional settings are crucial to understanding why floods would have a medium-term impact on population growth and urbanization. In the absence of flood protection and disaster relief, people in rural areas are more likely to leave in the medium-term.

In the next section, I investigate which factors could drive the difference in the impacts of a severe flood between rural and urban areas.

Table 5.5.1: Floods on log population - SSA

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days/100 | -2.7 | -3.6 | -3.8 | -5.7* |
| | (2.5) | (2.6) | (3.3) | (3.3) |
| Lagged flooded days/100 | | -5.9*** | | -7.7*** |
| | | (2.1) | | (2.4) |
| Flooded days/100 × Share urban population 1975 | | | 16.2* | 32.9*** |
| | | | (8.3) | (8.3) |
| Lagged flooded days/100 × Share urban population 1975 | | | | 18.8** |
| | | | | (7.6) |
| N | 361432 | 236426 | 361432 | 236426 |
| Within Rsquared | 0.0001 | 0.0005 | 0.0004 | 0.001 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

### 5.5.2 Nighttime light analysis

One of the main reasons the urban areas do not seem to experience the same out-migration as rural areas could be that they are better protected. While I can only observe the population every ten years, I can observe the change in nighttime lights yearly. I construct a yearly panel of night lights for every cell in my sample of Sub-Saharan countries and compare it with a measure of the number of flooded days. I still find an impact of floods on nightlights, but less persistent than the effect on population growth. The estimating equation for the next graph is the following: $Ln(NightLights_{c,t})$ night lights in cell $i$ at year $j$, Flooded Cell$_{c,j}$ a dummy if

cell $i$ has been flooded during year $j$, $\rho_i$ cell fixed effects and $\delta_j$ time fixed effects.

$$\begin{aligned} Ln(NightLights)_{c,j} = \alpha_0 \text{Flooded Cell}_{c,j} + \alpha_1 \text{Flooded Cell}_{c,j-1} + \\ \alpha_2 \text{Flooded Cell}_{c,j-2} + \alpha_3 \text{Flooded Cell}_{c,j-3} + \\ \alpha_4 \text{Flooded Cell}_{c,j-4} + \alpha_5 \text{Flooded Cell}_{c,j-5} + \\ + \rho_i + \delta_j + \epsilon_{c,j} \end{aligned}$$

I find that an increase in flooded days is associated with a decrease in night lights up to 5 years in the past in Figure 5.5.1. It indicates that urban areas were also impacted but might be recovering faster than rural areas. Unfortunately, I can not use night lights for the full sample as many areas are still not lit in Sub-Saharan Africa. The night light sample is thus more urban than the sample. The standard errors are larger at time 0 as the nighttime light uses a yearly average - meaning that time captures both before and just after the flood (there should be no impact of future floods on nighttime lights but a large impact just after).

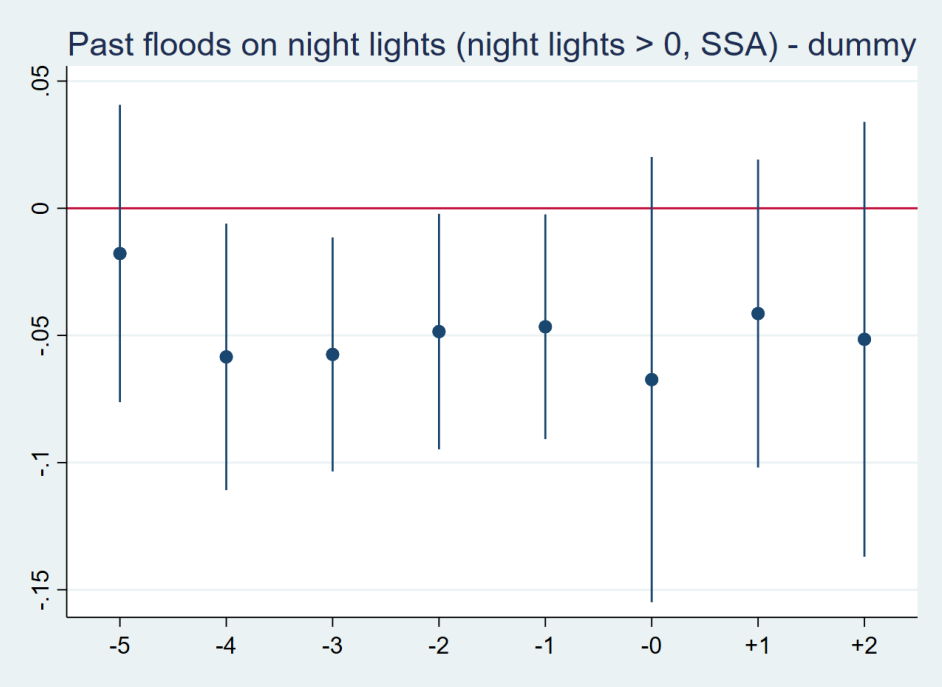
I reproduce the results from Table 5.5.1 in Table 5.A.9 but use Nighttime Lights instead of population. I find no negative impact in the most lit-up areas (columns 1 and 2). Once I introduce the interaction, I find a negative but not persistent impact in the rural areas and a positive but not persistent impact in the urban cells (columns 2 and 4).

These results are consistent with the finding of Kocornik-Mina et al. (2018), but the time effects are slightly more long-lasting in time. The sample includes much smaller cities, which might take longer to recover.

### 5.5.3 Demographic and health surveys analysis

I use Demographic and Health Surveys to shed light on the mechanisms that produce these differences between rural and urban areas. I find that severe floods are associated with worst outcomes: impacted households are on average poorer and suffer from a higher rate of malaria infection. However, these health and wealth effects are dissipated after a year. It suggests that the long-term differences in population growth rate emerging after large floods are more likely due to a tempering of out-migration in urban areas - either due to in-migration caused by lower housing prices or lower out-migration response.

Figure 5.5.1: Flooded days (yearly lags) on night lights growth




**5.5.3.1 Malaria infection** In this section, I analyze the impact of floods on health outcomes linked with flooding in the short and medium term. Malaria infection intensity has been linked to precipitations (Pascual et al. 2008; Krefis et al. 2011). Malaria has been shown to decrease economic growth (Gallup and Sachs 2001). I compare the probability of testing positive for malaria after a large flood in urban and rural areas using an event study type analysis for each individual $i$ in the DHS surveys. I run the following OLS and logit estimations:

$$Malaria_i = \alpha + \sum \beta_t^T Flooded_{i,t} + X_i\Phi + \epsilon_i$$

The outcome $Malaria_k$ is a dummy variable taking a value of 1 if the individual has tested positive for the malaria test. As a measure of floods, I use the number of months since the last flood. The base category is individuals that live in areas that have not been recorded as flooded in the last 18 months. The controls $X_i$ include age, gender, time of the interview and region dummy. The sample is all individuals in household DHS surveys in Sub-Saharan Africa with GPS coordinates. The regressions use the sampling weights provided in each survey and are resized using the sampling fraction of the country's population. All regressions are done separately in urban and rural clusters. Standard errors are clustered at the village level.

I run both OLS and logit regressions of time since the last flood on the probability of testing positive for malaria. I present the logit regression results in Table 5.A.10. In rural areas, I find an increased probability of testing positive up to a year after a flood on top of a much higher prevalence rate (the average rate of malaria is 35% in rural areas against 16% in urban areas). There is also a small peak in urban areas, but it is less spread in time.

I plot below in Figure 5.5.2 the coefficients of the OLS regressions. There is a clear increase in rural areas while the coefficients are close to zero in urban areas. It is consistent with a higher incidence rate in rural areas. However, it might also be due to better flood defences in urban areas.

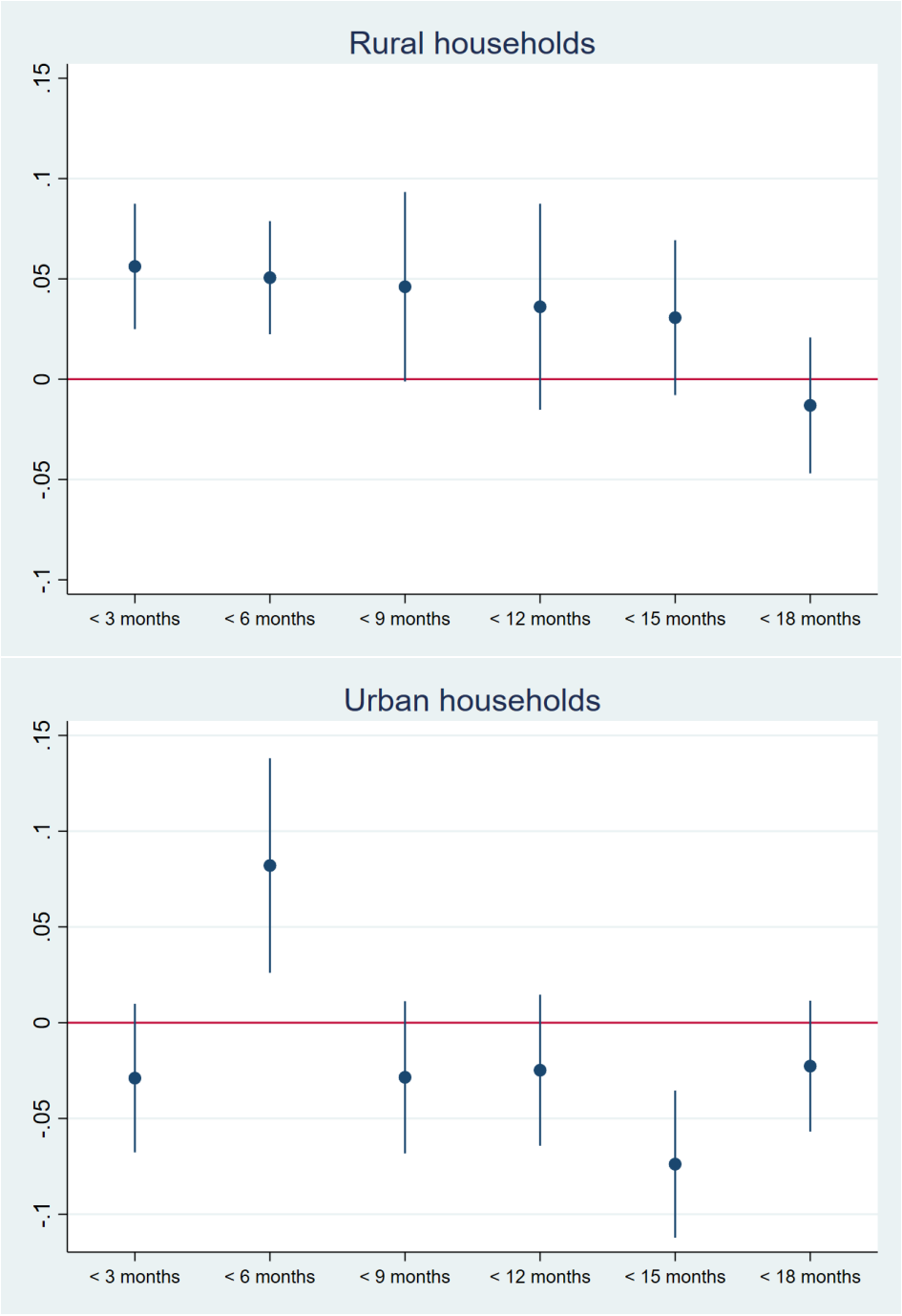
**5.5.3.2 Wealth effects** In the next section, I look at the short and medium-term effects of fluvial floods on households' wealth. The literature suggests that floods and extreme weather events are associated with diminishing resilience in the short-term (Jones, Ballon, and Engelhart 2018). I find that the more an area has suffered from severe floods, the more likely its inhabitants to fall in the lowest quantile of the wealth index in the DHS surveys. The DHS wealth index is a composite measure of households' living standards. The index is based on the following variables: the source of drinking water, type of toilet, sharing of toilet facilities, the material of the principal floor, walls, roof, cooking fuel, household services and possessions, such as electricity, TV, radio, watch, types of vehicles, agricultural land size owned, type and number of animals owned, bank account, types of windows. Floods could affect this measure by reducing the number of assets, using cheaper materials for housing construction, or destroying water access and sanitation facilities. Unfortunately, I do not observe households before and after a flood event. Still, I can capture the change in probability of being classified in the poorest quantile relative to the occurrence of floods.

I run the following event study estimation using an OLS and logit model using the same time since flood measure as in the previous malaria rate analysis:

$$Poorest_h = \alpha + \sum \beta_t^T Flooded_{h,t} + X_h\Phi + \epsilon_h$$

I use as a sample all households DHS surveys in Sub-Saharan Africa with GPS coordinates.

Figure 5.5.2: Probably of testing positive for malaria after a flood



The regressions use the sampling weights provided normalized by the sampling fraction of each survey.

The outcome $Poorest_h$ is a dummy variable taking a value of 1 if the household $h$ is classified as belonging to the poorest quantile according to a wealth index built on DHS data. All regressions are done separately in urban and rural clusters. I do an event study analysis by regressing being in the lowest wealth quantile on time in months since the last recorded flood. The base category is households that live in an area that has not been recorded as flooded in the past 18 months but has been in the last 10 years. I control for the number of household members, the number of children under 5, gender of the household head, age of the household head, month, year and country fixed effects.

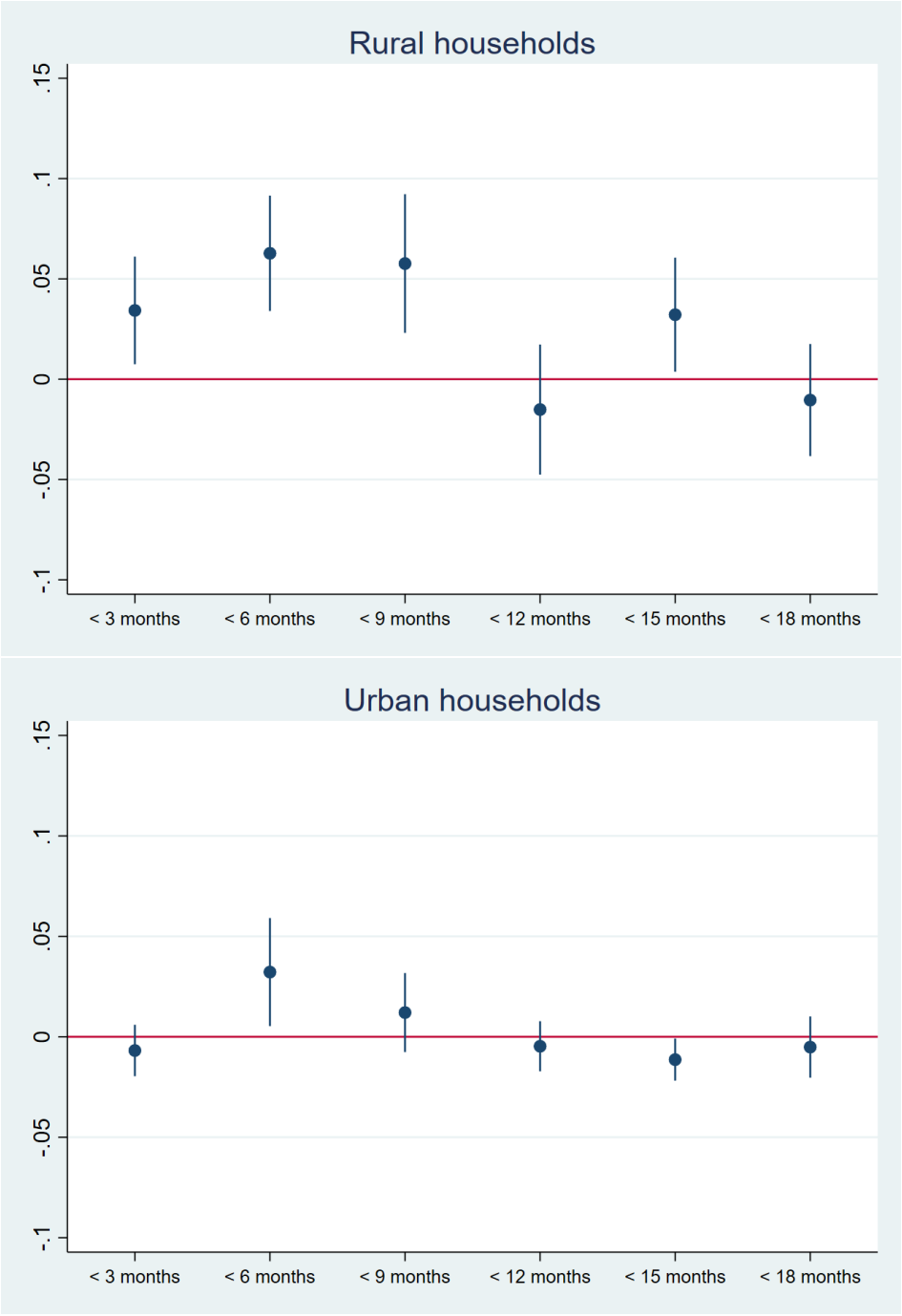
In Figure 5.5.3, I plot the OLS coefficients of the event study. Figure 5.5.3 shows an increase of about 7% in the probability of being categorized in the poorest quantile in the first 9 months after a flood in rural areas. The coefficient return to baseline after a year. The increase is much smaller in magnitude and shorter in time in urban areas. I include the corresponding logit regression in Table 5.A.11 in the Appendix.

Both health and wealth event studies show that floods have a short-term impact in rural areas. However, these effects are mostly dissipated after a year. To explain the differential growth rate in the medium-term, I look at the relationship between poverty and living in an area that has been repetitively flooded in the past 10 years. I run a similar regression to the previous event study but replace the main flood regressor with the count of floods in the last ten years. I show in Figure 5.5.4 how the probability of being in the lowest wealth category relates to the count of floods in the last ten years. The base category is households that live in an area with no recorded floods.

$$Poorest_h = \alpha + \beta CountFloods_h + X_h\Phi + \epsilon_h$$

I find a positive relationship between being in the lowest wealth quantile and the count of floods, especially in rural areas. The probability of being classified is increased by 15% in the

Figure 5.5.3: Probably of being in the lowest wealth quantile after a flood



most exposed rural areas. The coefficients in urban areas are all smaller than a 5% increase and barely significant. There is no strong relationship between poverty and the number of recorded floods in urban areas.

There are some caveats to the interpretation of these coefficients. The relationship is not necessarily causal as I can not observe migration patterns. Additionally, different channels could be at play in urban areas: households are wealthier and thus more resilient to shocks, and cities have better flood protection. In rural areas, I can not distinguish between an increase in poverty rate due to a decrease in agricultural output, destruction of assets, out-migration of richer households, or in-migration of poorer households. I do not find any persistent results in the wealth index, which points to sorting households in the medium term. As the population growth rate analysis shows a decrease in population growth in rural areas, it also seems to indicate that some households or working-age individuals leave the most impacted areas.

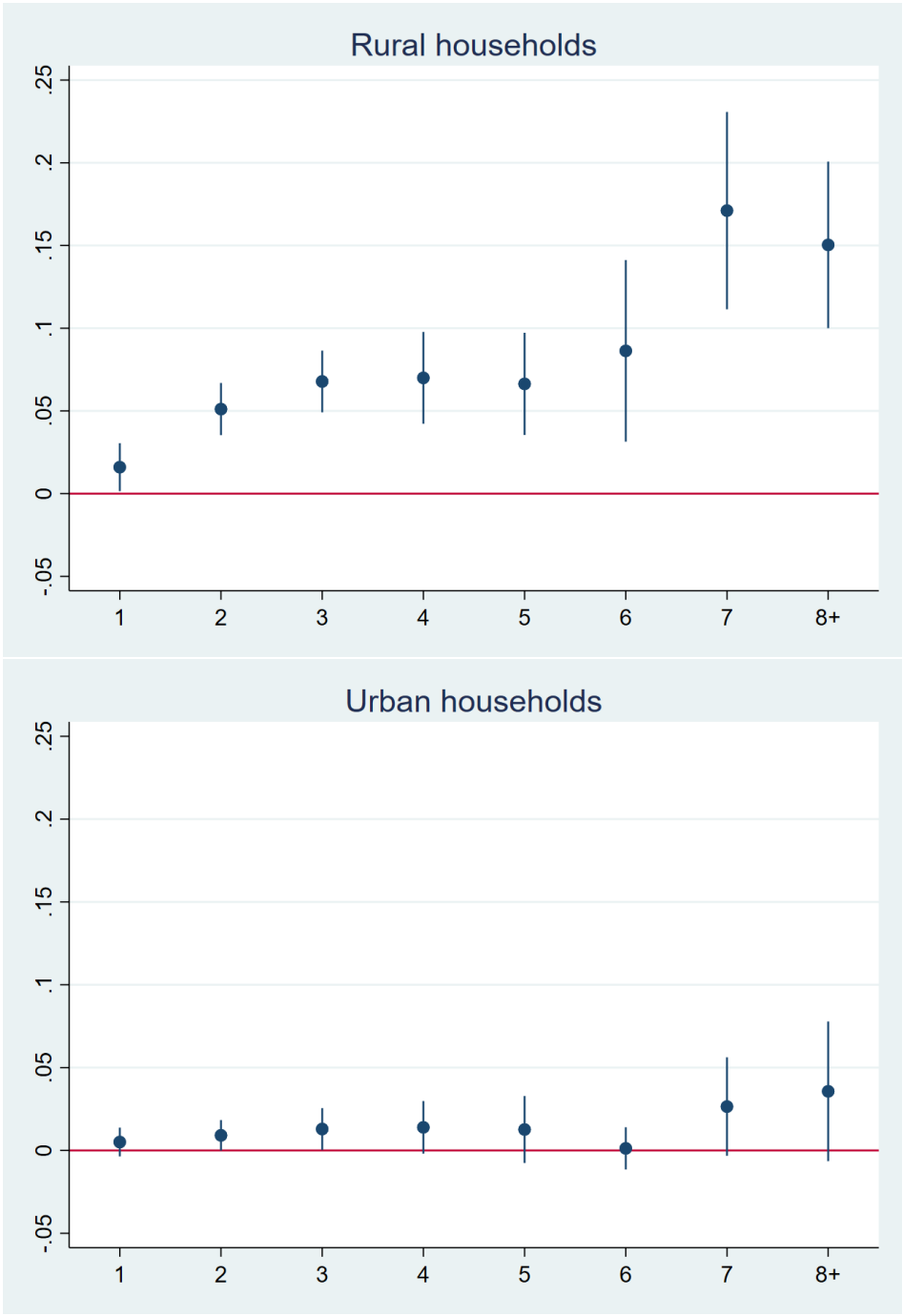
I do a placebo test by replacing the count of floods in the last ten years with the count of floods in the next ten years. I include the graphs in Appendix Figure 5.A.7. The small and positive relationship between the count of floods and poverty in urban areas disappears. However, the relationship between poverty and the count of floods in rural areas is still high. It indicates that the positive relationship in rural areas is probably driven by the sorting of poorer households in the most exposed areas and not only by impoverishment after a flood. I also include the logit results in Table 5.A.12.

## 5.6 Conclusion

In this paper, I explore the impacts of large floods on population growth in Sub-Saharan Africa. I pay particular attention to the difference between rural and urban areas.

I find that an increase in the number of severe floods is associated with a decrease in the population growth rate in rural areas in Sub-Saharan Africa over the 1990-2015 period. This finding is consistent with out-migration from flooded regions. However, this decrease is partially offset in more urban areas. Further analysis of the differences between rural and urban areas at the household level shows that they are both affected by increased poverty

Figure 5.5.4: Probably of being in the lowest wealth quantile after repetitive flooding



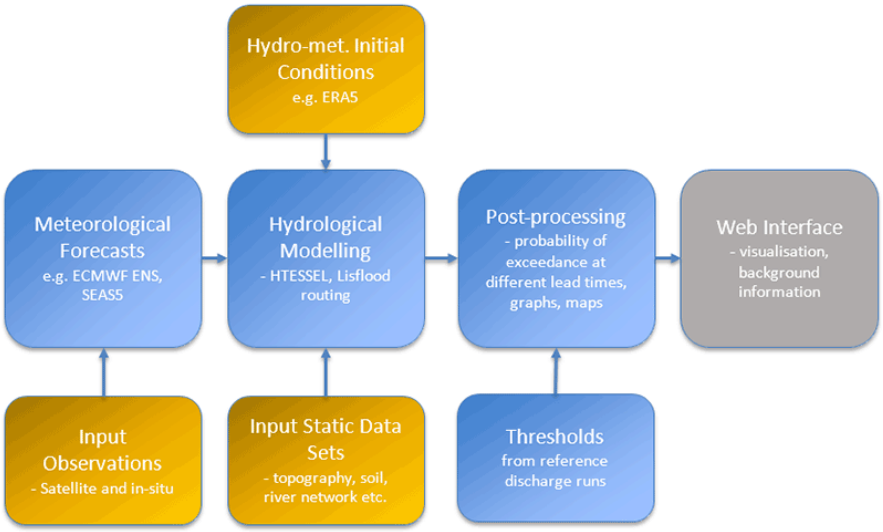
rates and disease incidence. However, nighttime light analysis seems to suggest that urban areas recover faster. Their base level of poverty and malaria incidence is also much lower. Finally, this work looks at the long-term costs of major natural catastrophes on local economic development. The poorest people are the most exposed to risks and the less able to cope. Repetitively flooded areas have the highest poverty rate. Improving flood management could lead to a long-term reduction in poverty and an increase in household resilience.

These findings are linked to the evidence that poor people are more exposed to floods and less able to cope with natural disasters. Nevertheless, a large number of people still choose to live in high flood-risk areas for economic and amenity-related reasons. I do not observe other economic variables such as wages or housing prices and thus can't derive welfare predictions from the results and comment on the optimality of people's locational choice. Similarly, I don't have a survey on people's flood risk perceptions, land prices, or short-term displacements. I only observe populations every ten years to fifteen years. The phenomenon that I observe is thus a medium-term decrease in the population growth rate in rural and less dense areas. It is different from a short-term negative shock: people being displaced or infrastructure being destroyed, followed by a short-term recovery - which would indicate that people do not update their flood risk perception and locate in dangerous places. I observe a medium to long-term effect of floods - meaning it is probably linked with the impacts of the floods listed in the literature: the attractivity of the places might decline because of the destruction of infrastructures, higher disease rate, decrease in households investments, decreased access to education for children, higher poverty rate (McMichael, Barnett, and McMichael 2012; Elsanousi et al. 2018; Lomnitz 2015).

# Appendix

## 5.A.1 GloFAS system

Figure 5.A.1: GloFAS system



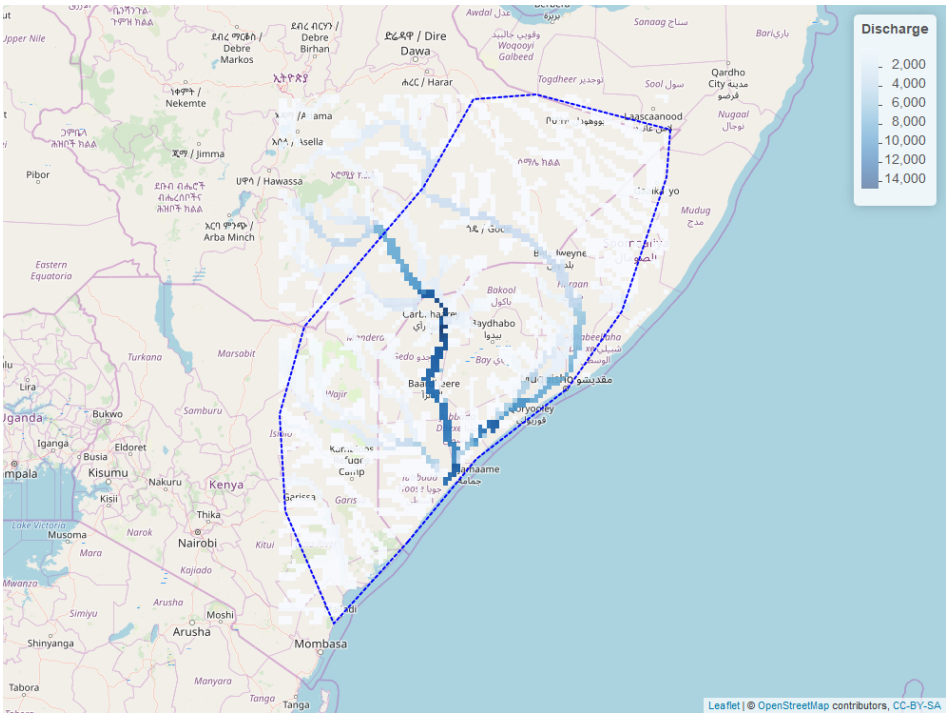

I present here the comparison of the GloFAS to the Dartmouth Flood Observatory (DFO) dataset by mapping individually each flood from the DFO dataset to the GloFAS grid data.

The Dartmouth Flood Observatory (DFO) dataset consists of about 4,900 floods collected from a "variety of news, governmental, instrumental, and remote sensing sources" worldwide. It covers 1985 to real-time, but the reporting improves dramatically over time. Shapefiles corresponding to the extent of the affected area are provided (the shape is usually larger than the actual flooded areas). The starting date, ending date, severity (on a scale from 1 to 2), death, displaced people, and costs are also recorded when available.

In Figure 5.A.2 and 5.A.3, I show how the two datasets relate using a flood in Somalia in 1997 as an example. First, I obtain the count of predicted flooded days for each cell by comparing the daily discharges to the 20-year threshold shown in Figure 5.A.4. I obtain a map of the count of flooded days for each cell during the duration of the flood in Figure 5.A.5. The grid cells in blue represent how many days are predicted as flooded by GloFAS during the period when the flood is considered "active" in the DFO dataset.

As shown in figure 5.A.5, the DFO reported flood (line in blue) encompasses a large area of "impacted" cells, including both areas reported flooded and areas affected by the floods (where the economic activity is disrupted by the flood). Only cells modelled in the GloFAS dataset can become flooded, so the extent is necessarily smaller than the DFO. To capture the impact of neighbouring floods, I also create for each cell a weighted average of the number of floods in neighbouring cells using a 30km radius[35].

Figure 5.A.2: River discharge day 1 of the flood, Somalia, 1997




### 5.A.2 GHSL

The Global Human Settlement Layer is a panel data on built-up land, population and degree of urbanization available for 1975, 1990, 2000, and 2014. It has been created by merging satellite imagery and census data.

Figure 5.A.3: River discharge day 41 of the flood, Somalia, 1997

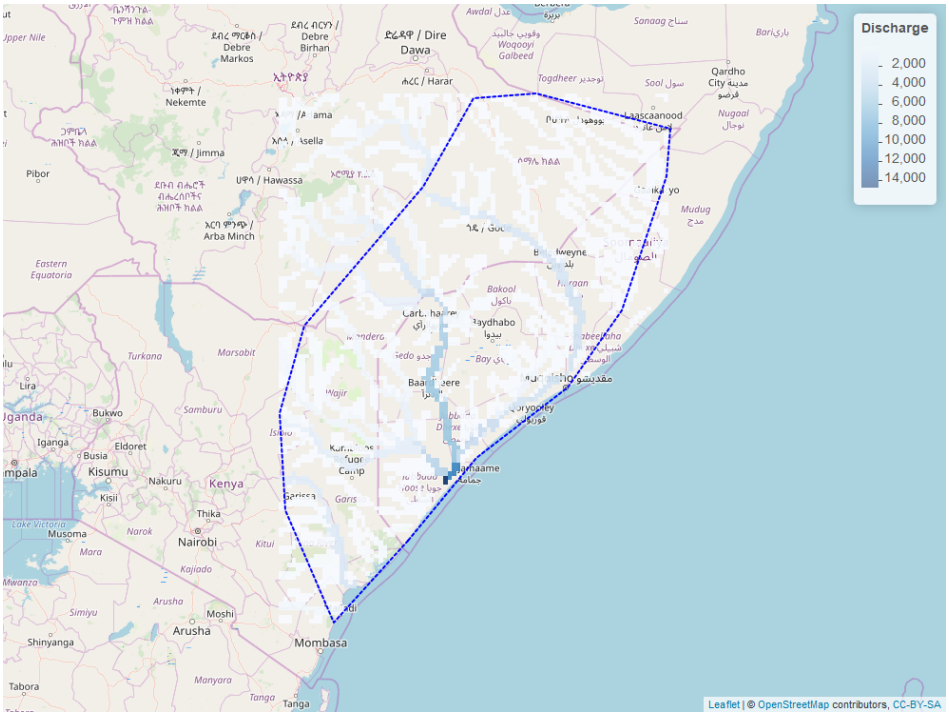



Figure 5.A.4: 1 in 20 return period threshold

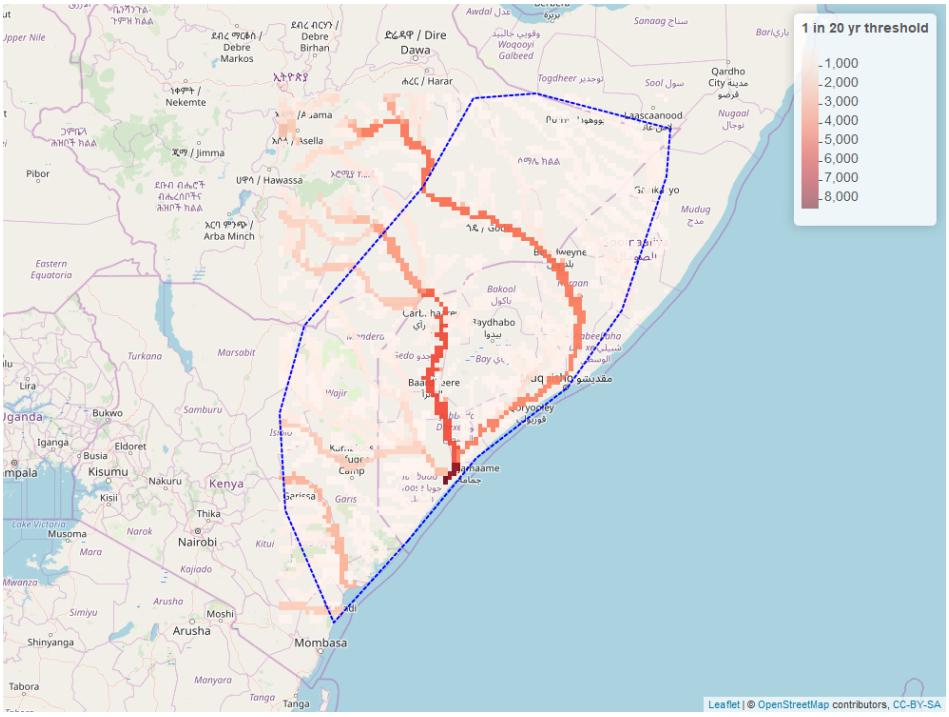


Figure 5.A.5: Number of flooded days (discharge > threshold), Somalia, 1997

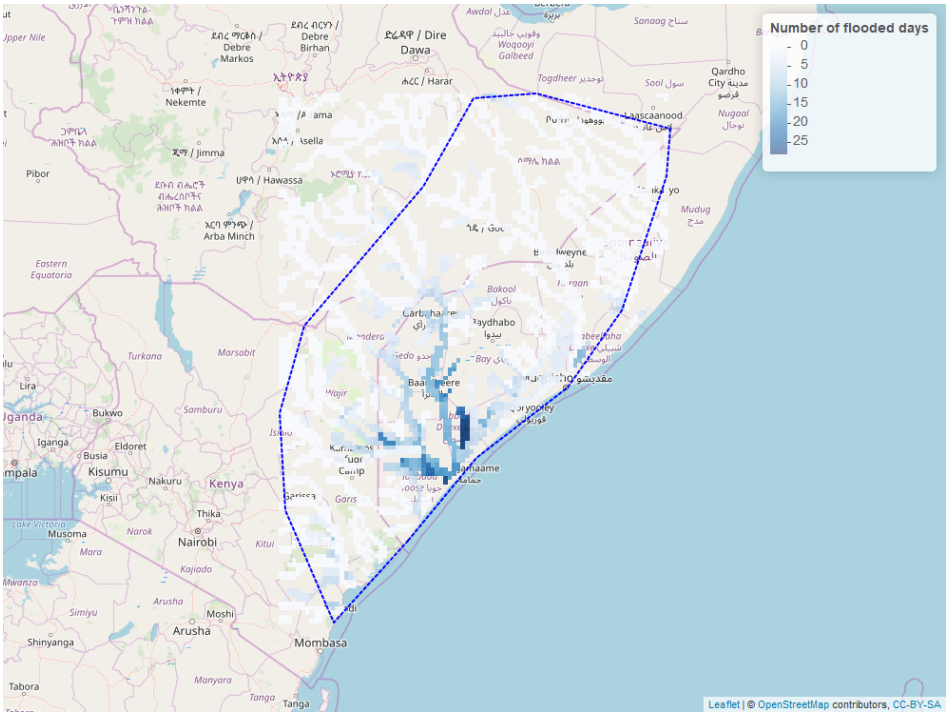



Figure 5.A.6: Conceptual schema of the GHSL input data, processing and products (Florczyk et al,. 2019)

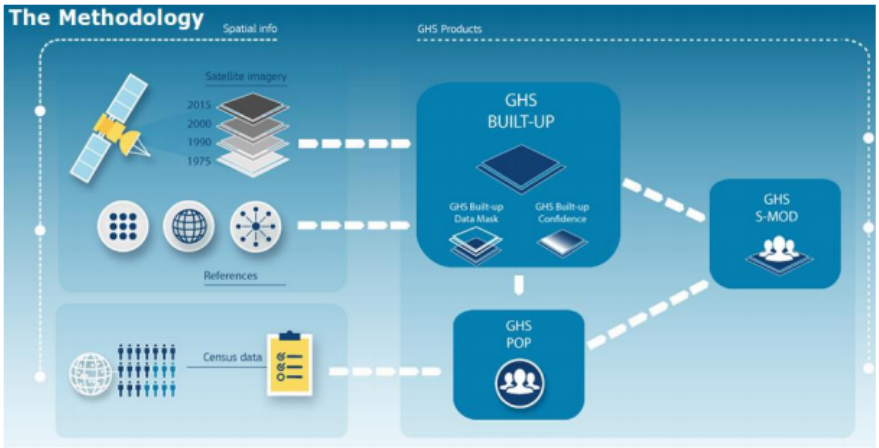


Table 5.A.1: Floods on log population by urban and rural sample separately

| | (1) Rural in 1975 | (2) Urban >0 in 1975 |
|---|---|---|
| Flooded days / 100 | -4.6 | 0.4 |
| | (3.2) | (3.7) |
| Lagged flooded days / 100 | -7.1*** | 0.2 |
| | (2.5) | (2.3) |
| N | 189608 | 46818 |
| Within Rsquared | 0.0006 | 0.000004 |
| Period Dummies | Yes | Yes |
| Country-Period Trend | Yes | Yes |
| FE | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: The first column corresponds to all cells in SSA without any urban population in 1975. The second column corresponds to any cells with some urban population in 1975.

### 5.A.3 Additional tables - population growth

Table 5.A.2: Floods on log population (urban population dummy)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | -2.7 | -3.6 | -4.0 | -5.9* |
| | (2.5) | (2.6) | (3.4) | (3.5) |
| Lagged flooded days / 100 | | -5.9*** | | -7.9*** |
| | | (2.1) | | (2.5) |
| Any urban in 1975 >0 × Flooded days / 100 | | | 13.7** | 25.4*** |
| | | | (6.7) | (6.5) |
| Any urban in 1975 >0 × Lagged flooded days / 100 | | | | 14.8** |
| | | | | (5.9) |
| N | 361432 | 236426 | 361432 | 236426 |
| Within Rsquared | 0.0001 | 0.0005 | 0.0004 | 0.001 |
| Period Dummies | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Any urban in 1975 is a dummy variable equal to 1 if the cell had any urban population in 1975 and 0 otherwise.

Table 5.A.3: Floods on log population (90th urban population percentile)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | -2.7 | -3.6 | -3.5 | -5.2* |
| | (2.5) | (2.6) | (3.1) | (3.0) |
| Lagged flooded days / 100 | | -5.9*** | | -7.3*** |
| | | (2.1) | | (2.2) |
| 90th pct. urban=1 × Flooded days / 100 | | | 13.8** | 28.7*** |
| | | | (6.5) | (6.6) |
| 90th pct. urban=1 × Lagged flooded days / 100 | | | | 16.2** |
| | | | | (6.2) |
| N | 361432 | 236426 | 361432 | 236426 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: 90th pct. urban is a dummy for cells with an urban population above the 90th percentile of the urban population distribution

Table 5.A.4: Floods on log population (urban population categorical)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | -2.7 | -3.6 | -4.0 | -5.9* |
| | (2.5) | (2.6) | (3.4) | (3.5) |
| Lagged flooded days / 100 | | -5.9*** | | -7.9*** |
| | | (2.1) | | (2.5) |
| Urban < 90th pct × Flooded days / 100 | | | 12.6* | 18.4** |
| | | | (6.6) | (7.4) |
| Urban < 90th pct× Lagged flooded days / 100 | | | | 11.4* |
| | | | | (6.2) |
| Urban > 90th pct × Flooded days / 100 | | | 14.3** | 29.5*** |
| | | | (6.9) | (6.9) |
| Urban > 90th pct × Lagged flooded days / 100 | | | | 16.8** |
| | | | | (6.5) |
| N | 361432 | 236426 | 361432 | 236426 |
| Within Rsquared | 0.0001 | 0.0005 | 0.0004 | 0.001 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Urban < 90th pct is a categorical variable for urban population under the 90th percentile. Urban > 90th pct is a categorical variable for urban population over the 90th percentile.

Table 5.A.5: Floods on log population (LIC)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | -6.3* | -7.1* | -8.6* | -11.6** |
| | (3.2) | (3.7) | (4.7) | (4.4) |
| Lagged flooded days / 100 | | -8.0*** | | -12.1*** |
| | | (1.2) | | (1.0) |
| Flooded days / 100 × Share urban in 1975 | | | 25.1** | 39.4*** |
| | | | (9.6) | (8.9) |
| Lagged flooded days / 100 × Share urban in 1975 | | | | 30.6*** |
| | | | | (6.5) |
| N | 218758 | 143970 | 218758 | 143970 |
| Within Rsquared | 0.0006 | 0.0009 | 0.001 | 0.003 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: The sample is any cell in a Low Income Country according to UN World Urban Prospect 2018

Table 5.A.6: Floods on log population (LMIC)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | 1.5 | 3.2** | 1.5 | 2.1 |
| | (0.9) | (1.5) | (1.1) | (1.9) |
| Lagged flooded days / 100 | | -0.5 | | -0.8 |
| | | (1.3) | | (1.6) |
| Flooded days / 100 × Share urban in 1975 | | | 0.2 | 8.3 |
| | | | (1.4) | (7.0) |
| Lagged flooded days / 100 × Share urban in 1975 | | | | 0.6 |
| | | | | (1.4) |
| N | 392313 | 257716 | 392313 | 257716 |
| Within Rsquared | 0.00005 | 0.0001 | 0.00005 | 0.0002 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: The sample is any cell in a Lower Middle Income Country according to UN World Urban Prospect 2018

Table 5.A.7: Floods on log population (UMIC)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | 0.06 | -1.7 | -2.5*** | -4.3** |
| | (0.5) | (1.8) | (0.8) | (2.1) |
| Lagged flooded days / 100 | | 1.3*** | | -0.5 |
| | | (0.3) | | (0.4) |
| Flooded days / 100 × Share urban in 1975 | | | 11.7*** | 17.4*** |
| | | | (1.5) | (3.8) |
| Lagged flooded days / 100 × Share urban in 1975 | | | | 7.7*** |
| | | | | (1.1) |
| N | 1027090 | 655480 | 1027090 | 655480 |
| Within Rsquared | 0.00000009 | 0.0001 | 0.0005 | 0.0007 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: The sample is any cell in a Upper Middle Income Country according to UN World Urban Prospect 2018

Table 5.A.8: Floods on log population (HIC)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days / 100 | 0.5 | -0.06 | 0.3 | -0.3 |
| | (0.9) | (1.4) | (1.0) | (1.7) |
| Lagged flooded days / 100 | | 1.2 | | 0.9 |
| | | (1.7) | | (2.1) |
| Flooded days / 100 × Share urban in 1975 | | | 2.2 | 2.4 |
| | | | (1.8) | (2.3) |
| Lagged flooded days / 100 × Share urban in 1975 | | | | 1.6 |
| | | | | (2.9) |
| N | 748931 | 488328 | 748931 | 488328 |
| Within Rsquared | 0.00002 | 0.0001 | 0.00005 | 0.0002 |
| Period FE | Yes | Yes | Yes | No |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: The sample is any cell in a High Income Country according to UN World Urban Prospect 2018

### 5.A.4 Additional tables - nighttime lights

Table 5.A.9: Floods on log population - SSA (Nightlight sample)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flooded days/100 | -2.2 | -3.7 | -4.4** | -6.4** |
| | (1.7) | (2.3) | (1.9) | (2.8) |
| Lagged flooded days/100 | | 0.4 | | -0.8 |
| | | (2.1) | | (2.7) |
| Flooded days/100 × Share urban 1975 | | | 12.6* | 24.6*** |
| | | | (6.4) | (8.2) |
| Lagged flooded days/100 × Share urban 1975 | | | | 10.1 |
| | | | | (7.0) |
| N | 137839 | 92116 | 137839 | 92116 |
| Within Rsquared | 0.00006 | 0.0002 | 0.0003 | 0.001 |
| Period FE | Yes | Yes | Yes | Yes |
| Country-Period Trend | Yes | Yes | Yes | Yes |
| Cell FE | Yes | Yes | Yes | Yes |

Standard errors clustered at country level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: The sample is any cell in SSA that has any light recorded at nighttime between 1992-2013 using NOAA's NGDC Earth Observation Group (EOG) yearly nighttime luminosity data Average Visible, Stable Lights, & Cloud Free Coverages.

35

### 5.A.5 Additional tables - DHS

Table 5.A.10: Probability of testing positive for malaria

| | (1) Urban | (2) Rural |
|---|---|---|
| Time since last flood | | |
| < 3 months | -0.155 | 0.283*** |
| | (0.157) | (0.0791) |
| < 6 months | 0.427*** | 0.253*** |
| | (0.157) | (0.0722) |
| < 9 months | -0.276 | 0.228** |
| | (0.199) | (0.116) |
| < 12 months | -0.286 | 0.205 |
| | (0.207) | (0.147) |
| < 15 months | -0.455*** | 0.155* |
| | (0.141) | (0.0938) |
| < 18 months | -0.228 | -0.0698 |
| | (0.144) | (0.0924) |
| Under 5 yrs old | 0.159 | 0.658*** |
| | (0.199) | (0.139) |
| Female | -0.0914** | -0.0284 |
| | (0.0384) | (0.0190) |
| N | 54871 | 151728 |
| Chi2 | 0.0882 | 0.133 |
| Pseudo-R2 | 1033.7 | 4169.0 |
| Year | Yes | Yes |
| Month | Yes | Yes |
| Country | Yes | Yes |

Standard errors clustered at the village level

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: Logit regression of time since last flood on testing positive for malaria at the individual level.

Time since last flood is a categorical variable divided into 3 months bins:

From less than 3 months since the last flood to less than 18 months ago.

The omitted base category is flooded more than 18 months ago.

Only households living in an area that has been flooded in the past 10 years are included.

Table 5.A.11: Probability of being in the lowest wealth quantile

| | (1) Urban | (2) Rural |
|---|---|---|
| Poorest quintile | | |
| < 3 months | -0.255 | 0.172*** |
| | (0.261) | (0.0656) |
| < 6 months | 0.621*** | 0.306*** |
| | (0.216) | (0.0680) |
| < 9 months | 0.361 | 0.277*** |
| | (0.274) | (0.0821) |
| < 12 months | -0.174 | -0.0760 |
| | (0.280) | (0.0856) |
| < 15 months | -0.481** | 0.161** |
| | (0.243) | (0.0717) |
| < 18 months | -0.180 | -0.0519 |
| | (0.283) | (0.0749) |
| HHs members | -0.0329*** | -0.0420*** |
| | (0.00988) | (0.00309) |
| Children under 5 | 0.288*** | 0.196*** |
| | (0.0231) | (0.00789) |
| Female head of HH | 0.0899* | 0.0191 |
| | (0.0538) | (0.0162) |
| Age of head of HH | 0.0179*** | 0.00570*** |
| | (0.00172) | (0.000459) |
| Month | Yes | Yes |
| Year | Yes | Yes |
| Country | Yes | Yes |
| N | 240789 | 540781 |
| Chi2 | 0.0871 | 0.0181 |
| Pseudo-R2 | 1750.2 | 1716.0 |

Standard errors clustered at village level

* $p < .1$, ** $p < .05$, *** $p < .01$

Figure 5.A.7: Probably of being in the lowest wealth quantile after repetitive flooding (placebo test)

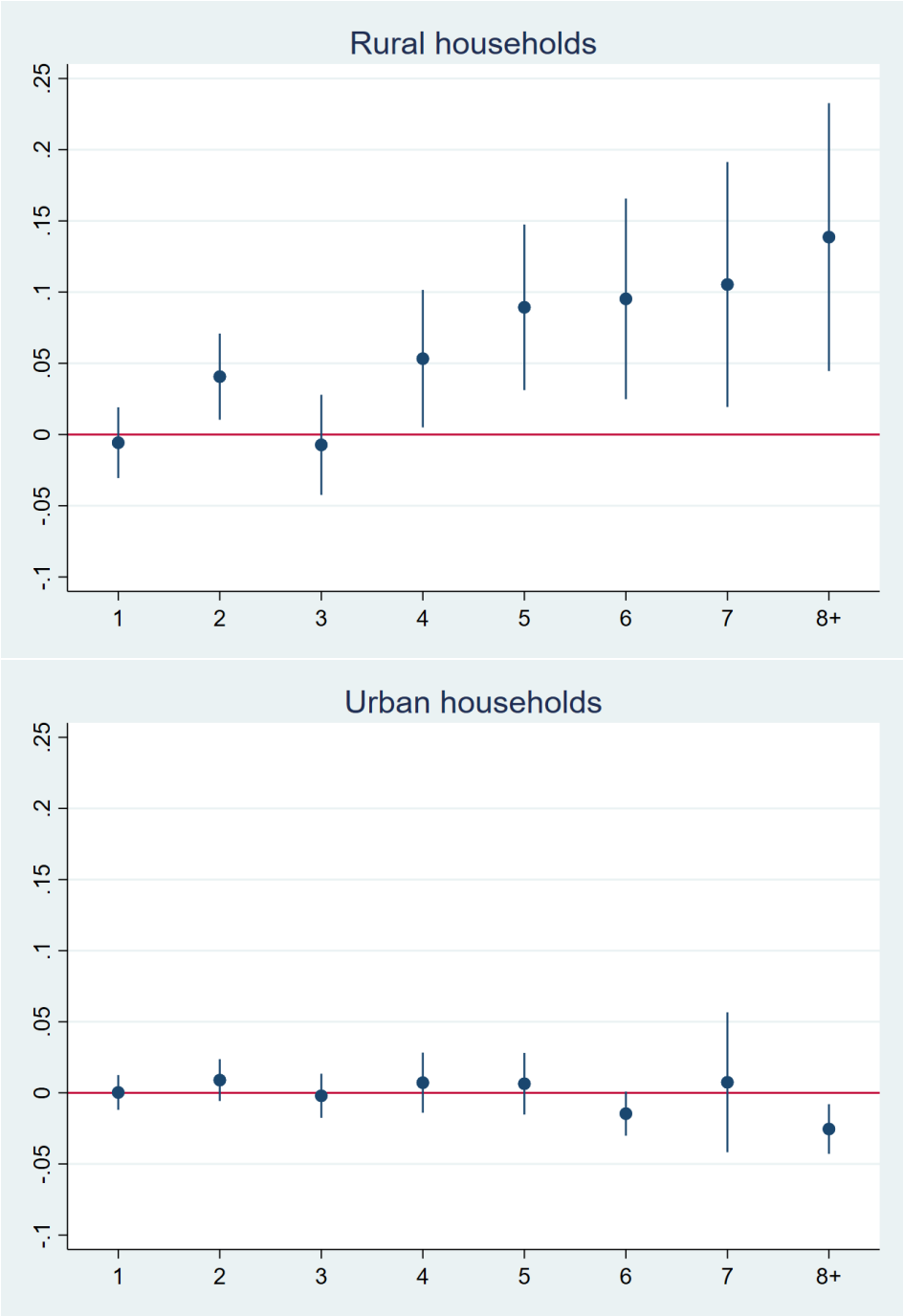


Table 5.A.12: Probability of being classified as poorest following repetitive flooding

| | Urban | | Rural | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Past 10 years | Next 10 years | Past 10 years | Next 10 years |
| Poorest quintile | | | | |
| 1 | 0.0460 | -0.0596 | 0.0904** | 0.0691* |
| | (0.147) | (0.158) | (0.0403) | (0.0370) |
| 2 | 0.349** | 0.103 | 0.246*** | 0.214*** |
| | (0.160) | (0.173) | (0.0436) | (0.0484) |
| 3 | 0.511*** | -0.00445 | 0.364*** | 0.0818 |
| | (0.193) | (0.203) | (0.0496) | (0.0603) |
| 4 | 0.717*** | 0.688*** | 0.421*** | 0.303*** |
| | (0.224) | (0.261) | (0.0705) | (0.0875) |
| 5 | 0.732*** | 0.168 | 0.328*** | 0.276** |
| | (0.264) | (0.365) | (0.0845) | (0.133) |
| 6 | 0.0699 | -1.433*** | 0.573*** | 0.574*** |
| | (0.281) | (0.452) | (0.134) | (0.172) |
| 7 | 1.071*** | 0.263 | 0.834*** | 0.221 |
| | (0.359) | (0.771) | (0.169) | (0.227) |
| 8+ | 1.144*** | -0.816 | 0.860*** | 0.693*** |
| | (0.422) | (0.956) | (0.132) | (0.266) |
| HHs members | -0.0514*** | -0.0481*** | -0.0464*** | -0.0457*** |
| | (0.0101) | (0.0102) | (0.00272) | (0.00272) |
| children under 5 | 0.299*** | 0.298*** | 0.198*** | 0.196*** |
| | (0.0220) | (0.0220) | (0.00692) | (0.00692) |
| Female head of HH | 0.0768 | 0.0795* | 0.0781*** | 0.0798*** |
| | (0.0479) | (0.0479) | (0.0142) | (0.0141) |
| Age of head of HH | 0.0150*** | 0.0149*** | 0.00617*** | 0.00606*** |
| | (0.00167) | (0.00167) | (0.000397) | (0.000395) |
| Month | Yes | Yes | Yes | Yes |
| Year | Yes | Yes | Yes | Yes |
| Country | Yes | Yes | Yes | Yes |
| N | 331721 | 330974 | 715585 | 714142 |
| Chi2 | 0.0844 | 0.0804 | 0.0188 | 0.0158 |
| Pseudo-R2 | 1550.1 | 1563.5 | 2014.6 | 1869.3 |

Standard errors clustered at the village level

* $p < .1$, ** $p < .05$, *** $p < .01$

## References


Achen, Christopher, and Larry Bartels. 2004. "Blind Retrospection: Electoral Responses to Droughts, Floods, and Shark Attacks." *Political Science*, 116–45. https://doi.org/10.1515/9781400888740-007.

Adger, W. Neil, Nigel W. Arnell, and Emma L. Tompkins. 2005. "Successful adaptation to climate change across scales." *Global Environmental Change* 15 (2): 77–86. https://doi.org/10.1016/J.GLOENVCHA.2004.12.005.

Albrecht, Frederike. 2021. "Natural hazards as political events: framing and politicisation of floods in the United Kingdom." *Https://Doi.org/10.1080/17477891.2021.1898926* 21 (1): 17–35. https://doi.org/10.1080/17477891.2021.1898926.

Aldred, Rachel, and John Dales. 2017. "Diversifying and normalising cycling in London, UK: An exploratory study on the influence of infrastructure." *Journal of Transport and Health* 4 (March): 348–62. https://doi.org/10.1016/J.JTH.2016.11.002.

Aldred, Rachel, Bridget Elliott, James Woodcock, and Anna Goodman. 2017. "Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age." *Transport Reviews* 37 (1): 29–55. https://doi.org/10.1080/01441647.2016.1200156/SUPPL_FILE/TTRV_A_1200156_SM7021.DOCX.

Aldred, Rachel, Anna Goodman, John Gulliver, and James Woodcock. 2018. "Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits." *Accident Analysis and Prevention* 117 (August): 75–84. https://doi.org/10.1016/j.aap.2018.03.003.

Alfieri, L, P Burek, E Dutra, B Krzeminski, D Muraro, J Thielen, and F Pappenberger. 2013. "GloFAS-global ensemble streamflow forecasting and flood early warning." *Hydrol. Earth Syst. Sci* 17: 1161–75. https://doi.org/10.5194/hess-17-1161-2013.

Arceneaux, Kevin. 2006. "The Federal Face of Voting: Are Elected Officials Held Accountable for the Functions Relevant to Their Office? on JSTOR." *Political Psychology* 27 (5): 731–54. https://www.jstor.org/stable/3792536#metadata_info_tab_contents.

Atkins, Graham, and Stuart Hoddinott. 2022. "Local government funding in England." https://www.instituteforgovernment.org.uk/explainers/local-government-funding-england.

Balboni, Clare. 2021. "In Harm's Way? Infrastructure Investments and the Persistence of

Coastal Cities."

Bates, Paul D, Matthew S Horritt, and Timothy J Fewtrell. n.d. "A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling." https://doi.org/10.1016/j.jhydrol.2010.03.027.

Baum-Snow, Nathaniel, and Lu Han. 2022. "The Microgeography of Housing Supply." https://drive.google.com/file/d/1WW8CwCxGjx9bvETbLZc8ERVo1Eb-h-Nx/view.

Bechtel, Michael M., and Jens Hainmueller. 2011. "How Lasting Is Voter Gratitude? An Analysis of the Short- and Long-Term Electoral Returns to Beneficial Policy." *American Journal of Political Science* 55 (4): 852–68. https://doi.org/10.1111/J.1540-5907.2011.00533.X.

Beltrán, Allan, David Maddison, and Robert Elliott. 2019. "The impact of flooding on property prices: A repeat-sales approach." *Journal of Environmental Economics and Management* 95 (May): 62–86. https://doi.org/10.1016/J.JEEM.2019.02.006.

Bernhofen, Mark V, Charlie Whyman, Mark A Trigg, P Andrew Sleigh, Andrew M Smith, Christopher C Sampson, Dai Yamazaki, et al. 2018. "A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique." *Environmental Research Letters* 13 (10): 104007. https://doi.org/10.1088/1748-9326/AAE014.

Besley, Timothy, and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *The Quarterly Journal of Economics* 117 (4): 1415–51. https://doi.org/10.1162/003355302320935061.

Bhuyan, Prajamitra, Emma J. McCoy, Haojie Li, and Daniel J. Graham. 2021. "Analysing the causal effect of London cycle superhighways on traffic congestion." *Https://Doi.org/10.1214/21-Aoas1450* 15 (4): 1999–2022. https://doi.org/10.1214/21-AOAS1450.

Bird, Julia, Piero Montebruno, and Tanner Regan. 2017. "Life in a slum: understanding living conditions in Nairobi's slums across time and space." *Oxford Review of Economic Policy* 33 (3): 496–520. https://doi.org/10.1093/oxrep/grx036.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and path dependence." *Quarterly Journal of Economics* 127 (2): 587–644. https://doi.org/10.1093/qje/qjs011.

Bohra-Mishra, Pratikshya, Michael Oppenheimer, and Solomon M Hsiang. 2014. "Nonlinear permanent migration response to climatic variations but minimal response to." *Source* 111

(27): 9780–85. https://doi.org/10.1073/pnas.1317166111.

Borusyak, Kirill, and Xavier Jaravel. 2018. "Revisiting Event Study Designs." *SSRN Electronic Journal*, March. https://doi.org/10.2139/ssrn.2826228.

Boustan, Leah Platt, Matthew E Kahn, Paul W Rhode, Leah Platt Boustan, Matthew E Kahn, and Paul W Rhode. 2012. "Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century." *American Economic Review* 102 (3): 238–44. https://doi.org/10.1257/aer.102.3.238.

Boustan, Leah Platt, Matthew E Kahn, Paul W Rhode, and Maria Lucia Yanguas. 2020. "The Effect of Natural Disasters on Economic Activity in US Counties: A Century of Data." *Journal of Urban Economics*. https://doi.org/10.3386/w23410.

Boyce, Ross, Raquel Reyes, Michael Matte, Moses Ntaro, Edgar Mulogo, Joshua P. Metlay, Lawrence Band, and Mark J. Siedner. 2016. "Severe Flooding and Malaria Transmission in the Western Ugandan Highlands: Implications for Disease Control in an Era of Global Climate Change." *Journal of Infectious Disease* 214 (9): 1403–10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5079365/.

Brakenridge, GR. 2016. "Global Active Archive of Large Flood Events, 1985-Present." https://floodobservatory.colorado.edu/Archives/index.html.

Brakenridge, Robert. 2019. "Dartmouth flood observatory." http://floodobservatory.colorado.edu/.

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner. 2006. "Causes of Sprawl: A Portrait from Space." *The Quarterly Journal of Economics* 121 (2): 587–633. https://doi.org/10.1162/QJEC.2006.121.2.587.

Castells-Quintana, David, Maria del Pilar Lopez-Uribe, and Tom McDermott. 2015. "Climate change and the geographical and institutional drivers of economic development," no. July: 1–50. http://www.lse.ac.uk/grantham.

Castells-Quintana, David, and Thomas K J Mcdermott. 2019. "Climate, Urbanization, and Conflict: The Effects of Weather Shocks and Floods on Urban Social Disorder (EWP 588)," no. 588. https://doi.org/10.22617/WPS190254-2.

Cavalcanti, F. 2018. "Voters sometimes provide the wrong incentives. The lesson of the Brazilian drought industry." *Munich Personal RePEc Archive*. https://www.semanticscholar.org/paper/fb83fc661fbc8d671638398d5bc7197a7308fe9a.

Cavallo, Eduardo A., and Ilan Noy. 2012. "The Economics of Natural Disasters: A Survey." *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.1817217.

Clugston, Harriet. 2021. "Climate change: councils in England spend almost £1.7bn fighting flooding and coastal erosion." https://www.nationalworld.com/news/environment/climate-change-councils-in-england-spend-almost-ps17bn-fighting-flooding-and-coastal-erosion-3341871.

Cohen, Emma. 2013. "Segregated bike lanes are safest for cyclists." CMAJ. https://doi.org/10.1503/cmaj.109-4468.

Cole, Shawn, Andrew Healy, and Eric Werker. 2012. "Do voters demand responsive governments? Evidence from Indian disaster relief." *Journal of Development Economics* 97 (2): 167–81. https://doi.org/10.1016/J.JDEVECO.2011.05.005.

Daniel, Vanessa E., Raymond J. G. M. Florax, and Piet Rietveld. 2009. "Flooding risk and housing values: An economic assessment of environmental hazard." *Ecological Economics* 69 (2): 355–65. https://doi.org/10.1016/j.ecolecon.2009.08.018.

Davis, Donald R, and David E Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review* 92 (5): 1269–89. https://doi.org/10.1257/000282802762024502.

De Chaisemartin, C., and X. D'Haultfoeuille. 2018. "Fuzzy differences-in-differences." *Review of Economic Studies* 85 (2): 999–1028. https://doi.org/10.1093/restud/rdx049.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2014. "What Do We Learn from the Weather ? The New Climate–Economy Literature." *Journal of Economic Literature* 52 (3): 740–98. https://doi.org/10.3386/w19578.

Department for Environment, Food & Rural Affairs (DEFRA). 2022. "Departmental Spending." https://www.parliament.uk/globalassets/documents/commons/scrutiny/defra-slides-2021-22.pdf#:$\sim$:text=How is DEFRA's investment spending changing in 2021-22%3F&text=the Environment Agency.,173 million from last year.

Department for Levelling Up, Housing, and Communities. 2022. "Local authority revenue expenditure and financing." https://www.gov.uk/government/collections/local-authority-revenue-expenditure-and-financing.

Deryugina, Tatyana. 2011. "The Dynamic Effects of Hurricanes in the US: The Role of Non-Disaster Transfer Payments." *MIT Center for Energy and Environmental Policy*

*Research*, no. May: 1–49. https://doi.org/10.1007/b101232.

———. 2017. "The fiscal cost of hurricanes: Disaster aid versus social insurance." *American Economic Journal: Economic Policy* 9 (3): 168–98. https://doi.org/10.1257/pol.20140296.

Desmet, Klaus, and J. Vernon Henderson. 2015. *The Geography of Development Within Countries*. Vol. 5. Elsevier B.V. https://doi.org/10.1016/B978-0-444-59531-7.00022-3.

Desmet, Klaus, Robert E. Kopp, Scott A. Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H. Strauss. 2021. "Evaluating the Economic Cost of Coastal Flooding." *American Economic Journal: Macroeconomics* 13 (2): 444–86. https://doi.org/10.1257/MAC.20180366.

Durand-Lasserve, Alain, Maÿlis Durand-Lasserve, and Harris Selod. 2015. "Land Delivery Systems in West African Cities : The Example of Bamako, Mali." *Land Delivery Systems in West African Cities: The Example of Bamako, Mali*, March. https://doi.org/10.1596/978-1-4648-0433-5.

Elsanousi, Yasir Elfatih Abdelrahim, Abbas Suleiman Elmahi, Irene Pereira, and Michel Debacker. 2018. "Impact of the 2013 Floods on the Incidence of Malaria in Almanagil Locality, Gezira State, Sudan." *PLoS Currents* 10 (October). https://doi.org/10.1371/currents.dis.8267b8917b47bc12ff3a712fe4589fe1.

Environment Agency. 2022a. "AIMS Spatial Flood Defences (inc. standardised attributes)." https://data.gov.uk/dataset/cc76738e-fc17-49f9-a216-977c61858dda/aims-spatial-flood-defences-inc-standardised-attributes.

———. 2022b. "Recorded Flood Outlines." https://data.gov.uk/dataset/16e32c53-35a6-4d54-a111-ca09031eaaaf/recorded-flood-outlines.

———. 2022c. "Risk of Flooding from Rivers and Sea." https://data.gov.uk/dataset/bad20199-6d39-4aad-8564-26a46778fd94/risk-of-flooding-from-rivers-and-sea.

Erman, Alvina, Elliot Motte, Radhika Goyal, Akosua Asare, Shinya Takamatsu, Xiaomeng Chen, Silvia Malgioglio, Alexander Skinner, Nobuo Yoshida, and Stephane Hallegatte. 2018. "The Road to Recovery The Role of Poverty in the Exposure, Vulnerability and Resilience to Floods in Accra Global Facility for Disaster Reduction and Recovery Social, Urban, Rural and Resilience Global Practice Poverty and Equity Global Practice." http://documents.worldbank.org/curated/en/301991534935351432/pdf/The-Road-to-Recovery-the-Role-of-Poverty-in-the-Exposure-Vulnerability-and-Resilience-to-Floods-in-

Accra.pdf http://www.worldbank.org/research.

Erman, Alvina, Mercedeh Tariverdi, Marguerite Obolensky, Xiaomeng Chen, Rose Camille Vincent, Silvia Malgioglio, Jun Rentschler, Stephane Hallegatte, and Nobuo Yoshida. 2019. "Wading Out the Storm." *Wading Out the Storm: The Role of Poverty in Exposure, Vulnerability and Resilience to Floods in Dar Es Salaam*, August. https://doi.org/10.1596/1813-9450-8976.

Fetzer, Thiemo. 2019. "Did Austerity Cause Brexit?" *American Economic Review* 109 (11): 3849–86. https://doi.org/10.1257/AER.20181164.

Fiorina, Morris P. 1978. "Economic Retrospective Voting in American National Elections: A Micro-Analysis." *American Journal of Political Science* 22 (2): 426. https://doi.org/10.2307/2110623.

Gallup, John Luke, and Jeffrey D. Sachs. 2001. "The Economic Burden of Malaria." https://www.ncbi.nlm.nih.gov/books/NBK2624/.

Ghimire, Ramesh, Susana Ferreira, and Jeffrey H. Dorfman. 2015. "Flood-induced displacement and civil conflict." *World Development* 66 (February): 614–28. https://doi.org/10.1016/j.worlddev.2014.09.021.

Glaeser, Edward L., Joseph Gyourko, and Raven E. Saks. 2006. "Urban growth and housing supply." *Journal of Economic Geography* 6 (1): 71–89. https://doi.org/10.1093/jeg/lbi003.

Goodman-Bacon, Andrew, and Jan Marcus. 2020. "Using Difference-in-Differences to Identify Causal Effects of COVID-19 Policies."

Green, Colin P, John S Heywood, and Maria Navarro. 2018. "Economics Working Paper Series Did the London Congestion Charge Reduce Pollution? Did the London Congestion Charge Reduce Pollution?" http://www.lancaster.ac.uk/lums/.

Hallegatte, Stephane, Adrien Vogt-Schilb, Mook Bangalore, and Julie Rozenberg. 2017. "Unbreakable." *Unbreakable: Building the Resilience of the Poor in the Face of Natural Disasters*. https://doi.org/10.1596/978-1-4648-1003-9.

Hallegatte, Stéphane, Adrien Vogt-Schilb, Julie Rozenberg, Mook Bangalore, and Chloé Beaudet. 2020b. "From Poverty to Disaster and Back: a Review of the Literature." *Economics of Disasters and Climate Change 2020 4:1* 4 (1): 223–47. https://doi.org/10.1007/S41885-020-00060-5.

———. 2020a. "From Poverty to Disaster and Back: a Review of the Literature." *Economics*

*of Disasters and Climate Change 2020 4:1* 4 (1): 223–47. https://doi.org/10.1007/S41885-020-00060-5.

Harari, Mariaflavia. 2020. "Cities in Bad Shape: Urban Geometry in India." *American Economic Review* 110 (8): 2377–2421. https://doi.org/10.1257/AER.20171673.

Healy, Andrew, and Neil Malhotra. 2009. "Myopic Voters and Natural Disaster Policy." *American Political Science Review* 103 (3): 387–406. https://doi.org/10.1017/S0003055409990104.

Herzog, Ian. 2020. "The City-Wide Effects of Tolling Downtown Drivers: Evidence from London's Congestion Charge." PhD thesis, University of Toronto. https://www.dropbox.com/s/j9w8g8qwqif1ide/HerzogLondonJMP.pdf?dl=0.

Hilber, Christian A. L., Teemu Lyytikäinen, and Wouter Vermeulen. 2011. "Capitalization of central government grants into local house prices: Panel data evidence from England." *Regional Science and Urban Economics* 41 (4): 394–406. https://doi.org/10.1016/J.REGSCIURBECO.2010.12.006.

Hochrainer, Stefan. 2009. *Assessing The Macroeconomic Impacts Of Natural Disasters: Are There Any?* Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-4968.

Hsiang, Solomon M, and Amir S Jina. 2014. "The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth: Evidence From 6,700 Cyclones." *National Bureau of Economic Research Working Paper Series* 20352: 1–70. https://doi.org/10.3386/w20352.

Husby, Trond, Henri L. F. de Groot, Marjan W. Hofkes, and Tatiana Filatova. 2018. "Flood protection and endogenous sorting of households: the role of credit constraints." *Mitigation and Adaptation Strategies for Global Change* 23 (2): 147–68. https://doi.org/10.1007/S11027-015-9667-7/TABLES/3.

IPCC. 2012. *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change*. Vol. 9781107025. https://doi.org/10.1017/CBO9781139177245.

Jedwab, Remi Camille, Edward Kerby, Alexander Moradi, Remi Jedwab, Edward Kerby, and Alexander Moradi. 2014. "History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya."

Jones, Lindsey, Paola Ballon, and Johannes von Engelhart. 2018. "How does resilience change

over time? Tracking post-disaster recovery suing mobile phone surveys."

Kahn, George a., and Scott Benolkin. 2007. "The role of money in monetary policy: why do the Fed and ECB see it so differently?" *Economic Review*, no. Q III: 5–36. http://ideas.repec.org/a/fip/fedker/y2007iqiiip5-36nv.92no.3.html.

Kahn, Matthew E. 2005. "The Death Toll from Natural Disasters: The Role of Income, Geography, and Institutions." *Review of Economics and Statistics* 87 (2): 271–84. https://doi.org/10.1162/0034653053970339.

Kaufmann, M, J Lewandowski, A Choryński, M Wiering, Maria Kaufmann, Jakub Lewandowski, Adam Choryński, and Mark Wiering. 2016. "Shock events and flood risk management: a media analysis of the institutional long-term effects of flood events in the Netherlands and Poland." *Ecology and Society, Published Online: Dec 19, 2016 | Doi:10.5751/ES-08764-210451* 21 (4). https://doi.org/10.5751/ES-08764-210451.

Keat Tang, Cheng, Christian Hilber, Diego Puga, Matthew Turner, Henry Overman, Steve Gibbons, Olmo Silva, Felipe Carozzi, and Sefi Roth and. 2016. "Traffic Externalities and Housing Prices: Evidence from the London Congestion Charge." Spatial Economics Research Centre, LSE. https://econpapers.repec.org/RePEc:cep:sercdp:0205.

Kellenberg, Derek K., and Ahmed Mushfiq Mobarak. 2008. "Does rising income increase or decrease damage risk from natural disasters?" *Journal of Urban Economics* 63 (3): 788–802. https://doi.org/10.1016/j.jue.2007.05.003.

Kocornik-Mina, Adriana, Thomas KJ McDermott, Guy Michaels, Ferdinand Rauch, David von Below, Vernon Henderson, Matthew Kahn, Somik Lall, Gianmarco León, and Steve Pischke. 2018. "Flooded cities." http://personal.lse.ac.uk/michaels/Kocornik-Mina_McDermott_Michaels_Rauch_Floods.pdf.

Koks, E. E., J. Rozenberg, C. Zorn, M. Tariverdi, M. Vousdoukas, S. A. Fraser, J. W. Hall, and S. Hallegatte. 2019. "A global multi-hazard risk analysis of road and railway infrastructure assets." *Nature Communications 2019 10:1* 10 (1): 1–11. https://doi.org/10.1038/s41467-019-10442-3.

Kousky, Carolyn, Erzo, F P Luttmer, Richard J Zeckhauser, C Kousky, E F P Luttmer, R J Zeckhauser, and John F Kennedy. 2006. "Private investment and government protection." *J Risk Uncertainty* 33: 73–100. https://doi.org/10.1007/s11166-006-0172-y.

Krefis, Anne Caroline, Norbert Georg Schwarz, Andreas Krüger, Julius Fobil, Bernard

Nkrumah, Samuel Acquah, Wibke Loag, et al. 2011. "Modeling the relationship between precipitation and malaria incidence in children from a holoendemic area in Ghana." *American Journal of Tropical Medicine and Hygiene* 84 (2): 285–91. https://doi.org/10.4269/ajtmh.2011.10-0381.

Kundzewicz, Zbigniew W, Shinjiro Kanae, Sonia I Seneviratne, John Handmer, Neville Nicholls, Pascal Peduzzi, Reinhard Mechler, et al. 2014. "Le risque d'inondation et les perspectives de changement climatique mondial et régional." *Hydrological Sciences Journal* 59 (1): 1–28. https://doi.org/10.1080/02626667.2013.857411.

Lall, Somik V., and Uwe Deichmann. 2012. "Density and Disasters : Economics of Urban Hazard Risk." *World Bank Research Observer* 27 (1): 74–105. https://doi.org/10.1093/WBRO/LKR006.

Lall, Somik V., Mattias K. A. Lundberg, and Zmarak Shalizi. 2008. "Implications of alternate policies on welfare of slum dwellers: Evidence from Pune, India." *Journal of Urban Economics* 63 (1): 56–73. https://doi.org/10.1016/J.JUE.2006.12.001.

Leape, Jonathan. 2006. "The London congestion charge." https://doi.org/10.1257/jep.20.4.157.

Li, Haojie, Hongliang Ding, Gang Ren, and Chengcheng Xu. 2018. "Effects of the London Cycle Superhighways on the usage of the London Cycle Hire." *Transportation Research Part A: Policy and Practice* 111 (May): 304–15. https://doi.org/10.1016/J.TRA.2018.03.020.

Li, Haojie, Daniel J. Graham, and Pan Liu. 2017. "Safety effects of the London cycle superhighways on cycle collisions." *Accident Analysis and Prevention* 99 (Pt A): 90–101. https://doi.org/10.1016/j.aap.2016.11.016.

Lin, Yatang, Thomas McDermott, and Guy Michaels. 2021. "Cities and the Sea Level." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3827610.

Lockwood, Benjamin, Francesco Porcelli, and James Rockey. 2022. "In the Grip of Whitehall? The Effects of Party Control on Local Fiscal Policy in the UK In the Grip of Whitehall? The Effects of Party Control on Local Fiscal Policy in the UK *." https://www.lgcplus.com/finance/exclusive-over-two-thirds-of-councils-to-raise-council-tax-by-.

Lomnitz, Cinna. 2015. "IFRC: World Disasters Report 2014: focus on Culture and Risk." *Natural Hazards* 77 (2): 1393–94. https://doi.org/10.1007/s11069-015-1655-4.

Marconcini, Mattia, Annekatrin Metz-Marconcini, Thomas Esch, and Noel Gorelick. 2021.

"Understanding Current Trends in Global Urbanisation-The World Settlement Footprint Suite." *GI_Forum 2021*. https://doi.org/10.1553/giscience2021_01_s33.

Masiero, Giuliano, and Michael Santarossa. 2020. "Earthquakes, grants, and public expenditure: How municipalities respond to natural disasters." *Journal of Regional Science* 60 (3): 481–516. https://doi.org/10.1111/JORS.12462.

McDermott, Thomas K. J. 2012. "Disasters and Development: Natural Disasters, Credit Constraints and Economic Growth." *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.1843494.

McDermott, Thomas K. J., Frank Barryy, and Richard S. J. Tol. 2014. "Disasters and development: Natural disasters, credit constraints, and economic growth." *Oxford Economic Papers* 66 (3): 750–73. https://doi.org/10.1093/oep/gpt034.

McMichael, Celia, Jon Barnett, and Anthony J. McMichael. 2012. "An III wind? Climate change, migration, and health." Public Health Services, US Dept of Health; Human Services. https://doi.org/10.1289/ehp.1104375.

Milly, P. C. D., R. T. Wetherald, K. A. Dunne, and T. L. Delworth. 2002. "Increasing risk of great floods in a changing climate." *Nature* 415 (6871): 514–17. https://doi.org/10.1038/415514a.

Moretti, Enrico. 2011. *Local labor markets*. https://doi.org/10.1016/S0169-7218(11)02412-9.

Muis, Sanne, Martin Verlaan, Hessel C. Winsemius, Jeroen C. J. H. Aerts, and Philip J. Ward. 2016. "A global reanalysis of storm surges and extreme sea levels." *Nature Communications 2016 7:1* 7 (1): 1–12. https://doi.org/10.1038/ncomms11969.

Mulvaney, Caroline A., Sherie Smith, Michael C. Watson, John Parkin, Carol Coupland, Philip Miller, Denise Kendrick, and Hugh Mcclintock. 2015. "Cycling infrastructure for reducing cycling injuries in cyclists." John Wiley; Sons Ltd. https://doi.org/10.1002/14651858.CD010415.pub2.

Narloch, Ulf, and Mook Bangalore. 2018. "The multifaceted relationship between environmental risks and poverty: new insights from Vietnam." *Environment and Development Economics* 23 (3): 298–327. https://doi.org/10.1017/S1355770X18000128.

Pascual, M., B. Cazelles, M. J. Bouma, L. F. Chaves, and K. Koelle. 2008. "Shifting patterns: Malaria dynamics and rainfall variability in an African highland." *Proceedings of the Royal Society B: Biological Sciences* 275 (1631): 123–32. https://doi.org/10.1098/rspb.2007.1068.

Peltzman, Sam. 1975. "The Effects of Automobile Safety Regulation." 4. Vol. 83. https://www.jstor.org/stable/pdf/1830396.pdf?refreqid=excelsior%3A368cd79968aab77e8e94315c44fe03be.

Pielke, Roger A, Joel Gratz, Christopher, W Landsea, Douglas Collins, Mark A Saunders, and Rade Musulin. 2008. "Normalized hurricane damage in the United States: 1900-2005." *Natural Hazards Review* 9 (1): 29–42. https://doi.org/10.1061/(ASCE)1527-6988(2008)9:1(29).

Priestley, Matthew D. K., Joaquim G. Pinto, Helen F. Dacre, and Len C. Shaffrey. 2017. "The role of cyclone clustering during the stormy winter of 2013/2014." *Weather* 72 (7): 187–92. https://doi.org/10.1002/WEA.3025.

Rahman, Muhammad Habibur, Nejat Anbarci, Prasad Sankar Bhattacharya, and Mehmet Ali Ulubaşoğlu. 2017. "Can extreme rainfall trigger democratic change? The role of flood-induced corruption." *Public Choice* 171 (3-4): 331–58. https://doi.org/10.1007/S11127-017-0440-1.

Rappaport, Jordan, and Jeffrey D. Sachs. 2003. "The United States as a Coastal Nation." *Journal of Economic Growth* 8 (1): 5–46. https://doi.org/10.1023/A:1022870216673.

Raschky, P. A. 2008. "Institutions and the losses from natural disasters." *Natural Hazards and Earth System Sciences* 8 (4): 627–34. https://doi.org/10.5194/nhess-8-627-2008.

Reif, Karlheinz, and Hermann Schmitt. 1980. "Nine second-order national elections: A conceptual framework for the analysis of European election results." *European Journal of Political Research* 8 (1): 3–44. https://doi.org/10.1111/J.1475-6765.1980.TB00737.X.

Rentschler, Jun, Paolo Avner, Mattia Marconcini, Rui Su, Emanuele Strano, Louise Bernard, Capucine Riom, and Stephane Hallegatte. 2022. "Rapid Urban Growth in Flood Zones : Global Evidence since 1985." *Policy Research Working Paper, World Bank Group*. http://documents.worldbank.org/curated/en/099546404212214703/IDU0ef8bc63a022b304b7c08e7c04aac815d4d98.

Rentschler, Jun, and Melda Salhab. 2020. "People in Harm's Way," Policy research working papers, October. https://doi.org/10.1596/1813-9450-9447.

Reynolds, Conor C. O., M. Anne Harris, Kay Teschke, Peter A. Cripton, and Meghan Winters. 2009. "The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature." Environ Health. https://doi.org/10.1186/1476-069X-8-47.

Richard Hornbeck, By, Daniel Keniston, Nava Ashraf, Leah Boustan, Leah Brooks, Bill

Collins, Brad DeLong, et al. 2017. "Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872 †." *American Economic Review* 107 (6): 1365–98. https://doi.org/10.1257/aer.20141707.

Ritchie, Hannah, and Max Roser. 2020. "Urbanization: Our World in Data." https://ourworldindata.org/urbanization#what-share-of-people-will-live-in-urban-areas-in-the-future https://ourworldindata.org/urbanization.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics* 125 (3): 1253–96. https://doi.org/10.1162/qjec.2010.125.3.1253.

Sampson, Christopher C., Andrew M. Smith, Paul B. Bates, Jeffrey C. Neal, Lorenzo Alfieri, and Jim E. Freer. 2015. "A high-resolution global flood hazard model." *Water Resources Research* 51 (9): 7358–81. https://doi.org/10.1002/2015WR016954.

Samson, J., D. Berteaux, B. J. Mcgill, and M. M. Humphries. 2011. "Geographic disparities and moral hazards in the predicted impacts of climate change on human populations." *Global Ecology and Biogeography* 20 (4): 532–44. https://doi.org/10.1111/j.1466-8238.2010.00632.x.

Scussolini, Paolo;, Jeroen C. J. H. Aerts, Brenden; Jongman, Laurens M; Bouwer, Hessel C. Winsemius, Hans De Moel, and Philip J. Ward. 2016. "FLOPROS: an evolving global database of flood protection standards." *Natural Hazards and Earth System Sciences* 16 (5): 1049–61. https://doi.org/10.5194/nhess-16-1049-2016.

Sun, Liyang, and Sarah Abraham. 2018. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, April. http://arxiv.org/abs/1804.05785.

Teale, Andrew. 2020. "Local Elections Archive Project." https://andrewteale.me.uk/leap/.

Tily, Geoff. 2020. "UK flooding – why years of austerity have left communities defenceless | TUC." https://www.tuc.org.uk/blogs/uk-flooding-why-years-austerity-have-left-communities-defenceless.

Toya, Hideki, and Mark Skidmore. 2007. "Economic development and the impacts of natural disasters." *Economics Letters* 94 (1): 20–25. https://doi.org/10.1016/j.econlet.2006.06.020.

Transport for London. 2014. "Finance and Policy Committee Item 5: Operational and Financial Performance Report and Investment Programme Report-Fourth Quarter, 2013/14."

———. 2018. "Cycling action plan."

Vousdoukas, Michalis I., Lorenzo Mentaschi, Evangelos Voukouvalas, Martin Verlaan, Svetlana Jevrejeva, Luke P. Jackson, and Luc Feyen. 2018. "Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard." *Nature Communications 2018 9:1* 9 (1): 1–12. https://doi.org/10.1038/s41467-018-04692-w.

Vousdoukas, Michalis I., Evangelos Voukouvalas, Alessandro Annunziato, Alessio Giardino, and Luc Feyen. 2016. "Projections of extreme storm surge levels along Europe." *Climate Dynamics* 47 (9-10): 3171–90. https://doi.org/10.1007/S00382-016-3019-5/FIGURES/12.

Wilson, Traci, and Sara Hobolt. n.d. "Allocating responsibility in multilevel government systems: voter and expert attributions in the European Union Article (Accepted version) (Refereed)." *European Union. The Journal of Politics*, no. 1: 102–13. https://doi.org/10.1086/678309.

Yamazaki, Dai, Daiki Ikeshima, Ryunosuke Tawatari, Tomohiro Yamaguchi, Fiachra O'Loughlin, Jeffery C. Neal, Christopher C. Sampson, Shinjiro Kanae, and Paul D. Bates. 2017. "A high-accuracy map of global terrain elevations." *Geophysical Research Letters* 44 (11): 5844–53. https://doi.org/10.1002/2017GL072874.