Keyi Fan
University of Toronto Econ &PCJ 24'
Harvard Kennedy School MPP 26'
Email: keyifan@hks.harvard.edu

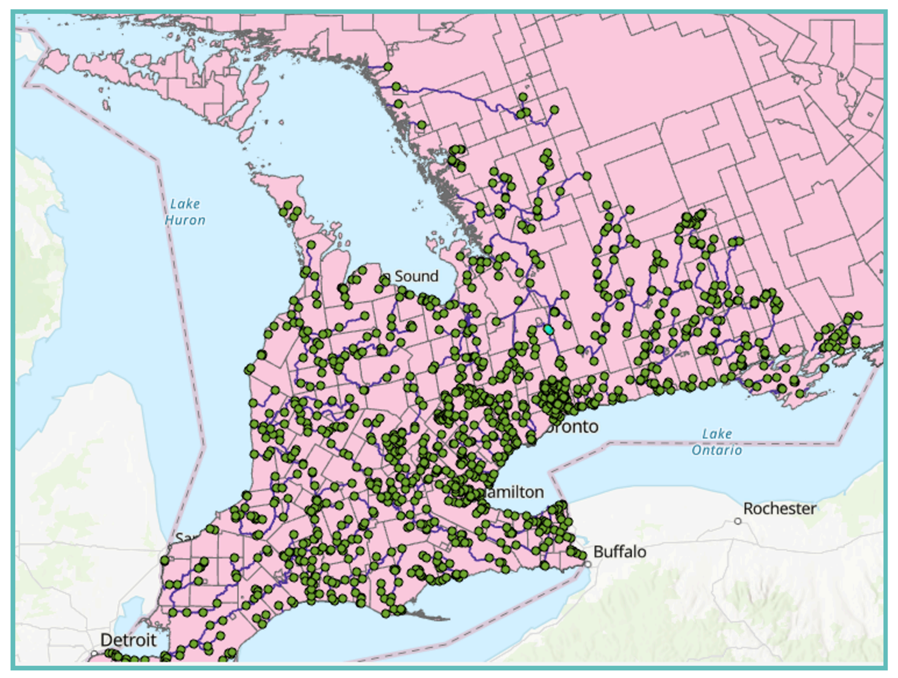
In the Ontario Greenbelt's Green-keeping : Evaluation of its Impact on Water Pollution and Environmental Conservation

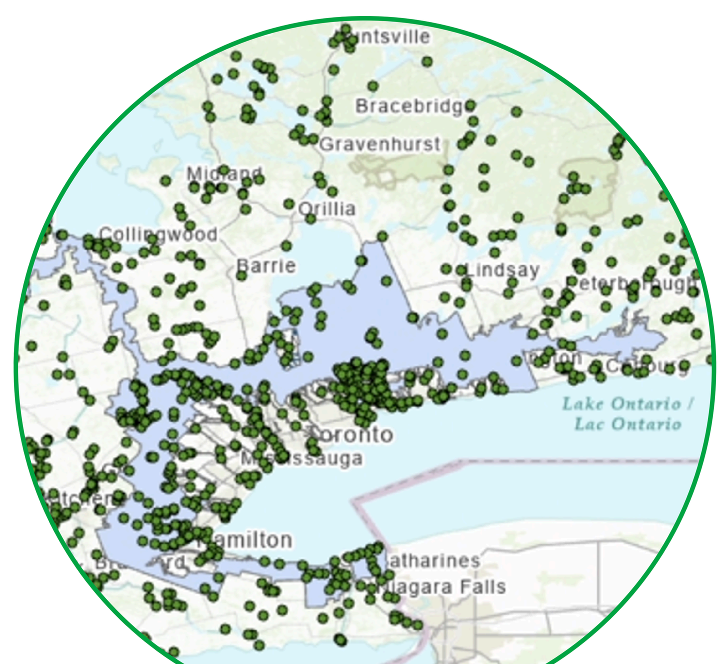
munk school OF GLOBAL AFFAIRS & PUBLIC POLICY
UNIVERSITY OF TORONTO

1. Introduction

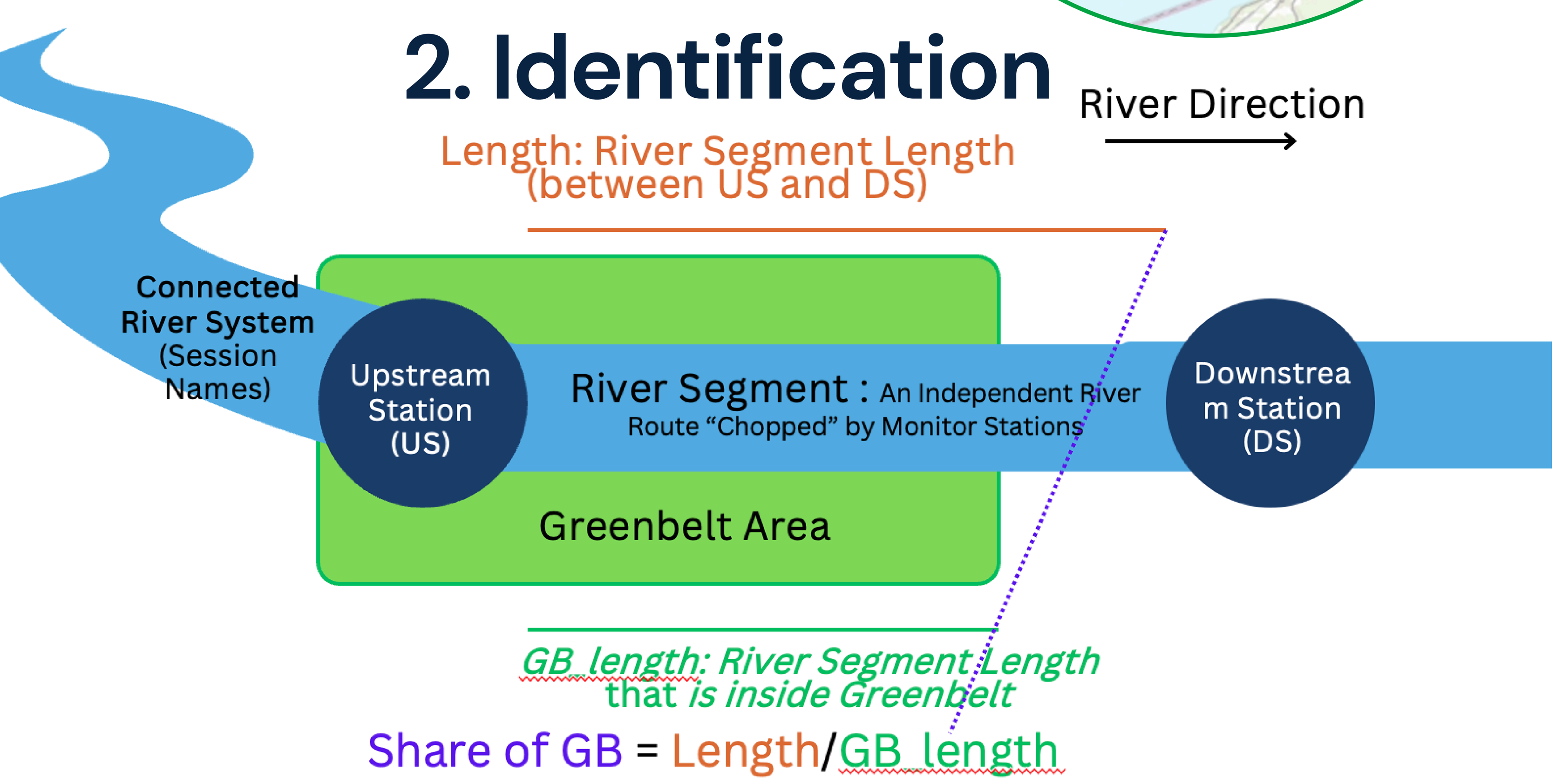
Research Question: -- "Did the Ontario Greenbelt Help Improve Southern Ontario Surface Water Quality?"

- The Ontario Greenbelt, established in 2005, is the world's largest protected greenbelt, encompassing 2 million acres across Southern Ontario. It aims to **curb urban sprawl, protect ecologically sensitive areas, and maintain the balance between human activities and the environment.**
- This study is **first to empirically examine the Ontario Greenbelt's impact** on surface water pollution using **causal inference** methods. It focus on key pollutants: Biological Oxygen Demand (BOD), Chromium, Lead, and Cadmium.
- Contributes **a novel dataset** combining high-resolution water quality monitoring data with detailed information on watercourse characteristics and census boundaries.
- I developed a **highly original algorithm** with ArcGIS & Python for processing **Canadian watercourse data**, contributing to the creation of similar datasets.





2. Identification



- I consider PWQMN monitoring stations along watercourses in Southern Ontario. For each pair, I identify upstream and downstream stations based **on flow direction,** and explore their **relative spatial relationship with the Greenbelt** (see Figure).
- By focusing on the change in **downstream minus upstream pollution before and after the Greenbelt (2005)**, I control for time-invariant differences between locations, isolating **ATE** .
- I assume the designation of the Greenbelt boundaries is exogenous to local water quality conditions. While the Greenbelt's overall location may be influenced by broad environmental and land use considerations, we argue **that the precise delineation of its boundaries is unlikely to be systematically related to pre-existing differences**
- Additionally, I exploit the spatial discontinuity in the Greenbelt's coverage for surrounding rivers and control for water system affiliation of each pair of stations.
- The corresponding **DiD model** is below :

$$\Delta \ln(BOD)_{i,t} = \alpha + \beta_1 \text{GB share}_{i,t} + \beta_2 \text{Post GB}_{i,t} + \beta_3 (\text{GB Share} \times \text{Post Gb})_{i,t} + \beta_4 \text{River Length}_i + \beta_5 \text{CSD}_i + \lambda_i \text{SessionName}_i + \gamma_i \text{SubrouteID}_i + \mu_t \text{Year}_t + \phi_s \times \text{Month}_{i,t} + \varepsilon_{i,t}$$

3. Data

- Data sources: Provincial (Stream) Water Quality Monitoring Network (**PWQMN**) data from 2000-2020, **Ontario Integrated Hydrology Data**, and **Census Boundary Files.**
- PWQMN provides water quality measurements for BOD, Chromium, Lead, and Cadmium at monitoring stations across Ontario (see summary statistics below).
- Ontario Integrated Hydrology Data identifies **watercourses intersecting monitoring stations in PWQMN and upstream/downstream relationships.**
- Census Boundary Files link stations to census subdivisions, controlling for **socioeconomic** factors.

Summary statistics are below for BOD pollutants:

	Control					Treatment				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
US Pollution (mg/L)	906	1.32	1.37	.15	14	3156	1.49	1.55	.2	18.2
DS Pollution (mg/L)	906	1.39	1.52	.1	20.2	3156	1.48	1.54	.1	18.2
Differences in Log(Pollutions)	906	0.07	0.68	-2.833213	3.828641	3156	-0.00	0.80	-3.367296	4.043051
Log(Pollution_US)	906	-0.04	0.76	-1.89712	2.639057	3156	0.08	0.78	-1.609438	2.901422
Log(Pollution_DS)	906	0.04	0.72	-2.302585	3.005683	3156	0.08	0.78	-2.302585	2.901422
Post GB	906	0.20	0.40	0	1	3156	0.34	0.47	0	1
Share of River in GB	906	0.00	0.00	0	0	3156	1.00	0.04	.0593062	1
length of River in GB	906	0.00	0.00	0	0	3156	2613.16	1297.78	148.3676	5640.176
River Length	906	3204.74	1298.44	1338.018	5957.169	3156	2619.09	1294.88	1030.809	5640.176
Observations	906					3156				

4. Results

Table 5: Greenbelt Policy Impact on Pollutants

	(1) BOD Basic	(2) BOD FE	(3) Chromium Basic	(4) Chromium FE	(5) Lead Basic	(6) Lead FE	(7) Cadmium Basic	(8) Cadmium FE
Share of GB	-1.590***	-4.698***	0.0220	4.717***	0.00961	4.723***	0.0131	4.738***
	(0.0555)	(1.21e-10)	(0.0349)	(0.0130)	(0.0282)	(0.0114)	(0.0170)	(0.0133)
Post GB=1	0.276***	0.308***	-0.630	-0.726	-0.926***	-1.022**	-0.884*	-0.908*
	(0.0531)	(0.0362)	(0.337)	(0.371)	(0.281)	(0.315)	(0.428)	(0.449)
Post GB=1 × Share of GB	-0.124***	-0.147***	0.162	0.222	0.00692	-0.00602	-0.0523	-0.0190
	(0.0121)	(0.0101)	(0.295)	(0.336)	(0.194)	(0.235)	(0.219)	(0.266)
River Length	0.0000953***	-0.000272***	0.0000837***	-0.0000942***	0.000104*	-0.0000895***	0.0000831**	-0.0000783***
	(0.000000997)	(1.49e-14)	(0.0000226)	(0.00000949)	(0.0000417)	(0.00000833)	(0.0000283)	(0.00000976)
Observations	3,942	4,018	2,172	2,172	1,397	1,397	2,241	2,241
R-squared Overall	0.158	0.341	0.074	0.202	0.074	0.114	0.079	0.101
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subroute Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses
Standard errors clustered at the water session level (us_sessionname) are reported in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

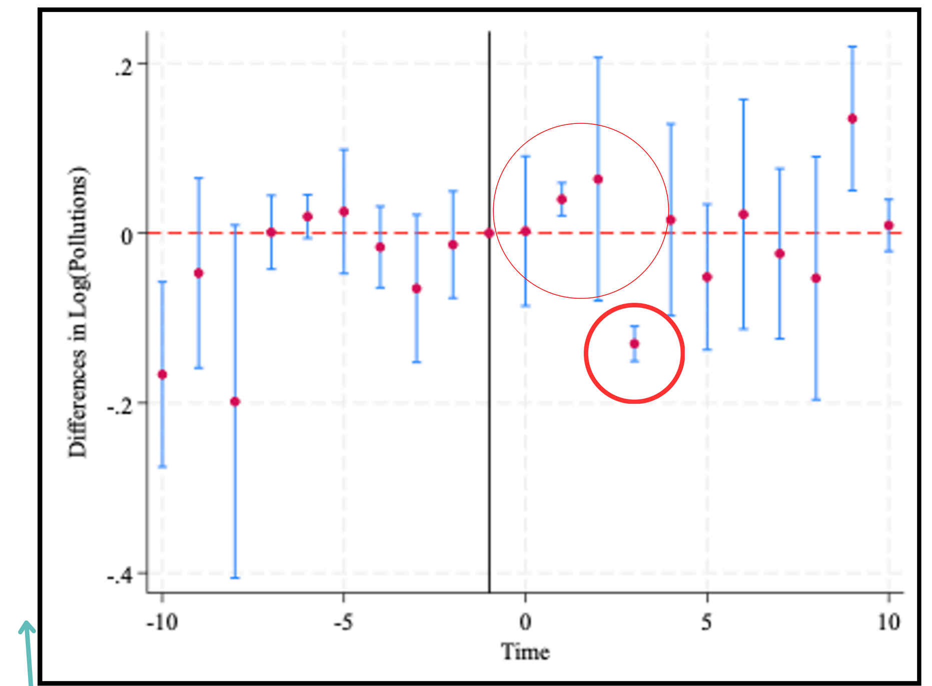
Placebo Tests

- The Greenbelt **significantly reduced BOD levels** in river segments with a higher share of the protected area (Columns 1-2).
  - A 100 percentage point increase in Greenbelt share(that is, the whole segment is in the greenbelt) is associated with a **12-15% decrease in BOD** after implementation, holding other factors constant.
- Placebo tests using heavy metal pollutants (Chromium, Lead, Cadmium) **show no consistent or significant effects** (Columns 3-8), supporting the causal interpretation.
- The **event study analysis** estimated using the model below *(see result in figure on the left)* suggests **a persistent BOD reduction**, with treatment effects stabilizing post-Greenbelt.

$$\Delta \ln(BOD)_{i,t} = \alpha + \sum_{j=2000}^{2020} \beta_j (\text{GBShare} \times \text{Year})i,t + \beta_4 \text{River Length}i + \beta_5 \text{CSD}i + \lambda_i \text{SessionName}i + \gamma_i \text{SubrouteID}i + \mu_t \text{Year}t + \phi_s \times \text{Month}i,t + \varepsilon i,t \quad (2)$$

- While not perfect, the event study provides some evidence supporting **the parallel trends assumption --** the pollution level between control and treatment were broadly similar before the Greenbelt's implementation.
- The treatment effect appears to **manifest around 3 years after the Greenbelt's announcement.** This delay could be attributed to the time needed for the policy to take effect and influence land use patterns. This finding is consistent with the expectation that the **Greenbelt's impact on water quality would emerge gradually.**

5. Robustness



- To ensure the reliability of my findings, I am in the process of finalizing and reporting a series of robustness checks.
- So far, the results from experiments are **robust to alternative specifications**, including different fixed effects and clustering of standard errors at various levels.
- I test for parallel trends in the pre-Greenbelt period using an event study design, which broadly **supports the identifying assumption.**
- Placebo tests using heavy metal pollutants further strengthen the **credibility of the causal interpretation.**
- The consistency of the findings across different specifications and tests reinforces the validity of the estimated treatment effects.

6. Key Findings

- The Ontario Greenbelt has been effective in **reducing organic pollution, as measured by BOD**, in the river segments within its boundaries.
- The estimated treatment effects are **substantial, statistically significant, and robust** to various specifications and sensitivity tests.
- The results highlight the potential of **land use planning policies**, such as greenbelts and urban growth boundaries, to generate significant environmental benefits.
- The findings contribute to the growing body of literature on the effectiveness of **urban containment policies** in promoting sustainable development and mitigating environmental challenges.
- The study demonstrates the value of quasi-experimental methods and high-resolution spatial data in evaluating the **causal impacts** of environmental policies.

7. Next Steps

- **Collaborate with municipalities and environmental science experts** to interpret the results and assess their **implications** for local water resource management.
  - Importantly, a more comprehensive explanation for results on heavy metal will be provided.
  - I expect **stronger qualitative evidence** to support causal claim.
- Explore the interactions between the Greenbelt and **other drivers of water quality**, such as land use **practices**, **wastewater** treatment, highway/**construction waste** and **agricultural** activities.
- Investigate the potential **treatment heterogeneity** in the Greenbelt's impact across different regions, watersheds, or land use types.
- Compare the **cost-effectiveness** and distributional consequences of the Greenbelt to alternative policy instruments for reducing water pollution.
- Extend the analysis to assess the Greenbelt's broader environmental, social, and economic impacts, such as its effects on **biodiversity**, **recreational opportunities**, and **farm productivity.**

Acknowledgement

I want to express my gratitude to my two direct supervisors, Prof. Gustavo Bobonis and Prof. Eduardo Souza Rodrigues. Also, I want to thank Prof. Arthur Blouin, Prof. Moussab, Prof. Paola Salardi, Prof. Courtney Ward, Sean Elliott and Grant Benjamin for valuable guidance during the creation of my very first metrics paper. Any remaining errors are solely my responsibilities.I also invite you to **vote for this paper as the Best Undergraduate Paper and Best Presentation.**