Canadian Journal of Regional Science
Revue canadienne des sciences régionales

Using Google Data to Estimate the Effects of Regional Mobility on Daily COVID-19 Cases: Evidence from Ontario Public Health Units

Anindya Sen

Volume 44, Number 3, 2021

URI: https://id.erudit.org/iderudit/1086217ar
DOI: https://doi.org/10.7202/1086217ar

See table of contents

Publisher(s)

Canadian Regional Science Association / Association canadienne des sciences régionales

ISSN

0705-4580 (print)
1925-2218 (digital)

Explore this journal

Cite this article

Sen, A. (2021). Using Google Data to Estimate the Effects of Regional Mobility on Daily COVID-19 Cases: Evidence from Ontario Public Health Units. *Canadian Journal of Regional Science / Revue canadienne des sciences régionales*, *44*(3), 167–175. https://doi.org/10.7202/1086217ar

Article abstract

From September 2020 onwards, COVID-19 cases have rapidly increased across Canada. This study estimates the effects of Google population mobility indicators on daily COVID-19 cases to evaluate the impacts of public movements across different regions in Ontario. We focus on Ontario as Google mobility data are available for Public Health Units (PHUs) for that province. Results based on pooled data from May 1st – November 15th imply that higher mobility at retail stores is significantly correlated with an increase in daily COVID-19 cases. However, empirical estimates from individual PHU level time-series models reveal regional differences, as these findings are based primarily on the relationship between retail mobility and daily cases for the Public Health Units (PHUs) of Toronto and Peel. These results support the implementation of region-specific lockdowns. Further, different specifications generate daily COVID-19 forecasts for Peel and Toronto that are on average, approximately 6%-9% different from actual values. The models of this research should be of value to local health authorities who are in search of simple models that are not computationally intensive and are capable of generating reliable forecasts for specific regions.

CANADIAN JOURNAL OF REGIONAL SCIENCE
REVUE CANADIENNE DES SCIENCES RÉGIONALES

CANADIAN REGIONAL SCIENCE ASSOCIATION
ASSOCIATION CANADIENNE DE SCIENCE RÉGIONALE

USING GOOGLE DATA TO ESTIMATE THE EFFECTS OF REGIONAL MOBILITY ON DAILY COVID-19 CASES: EVIDENCE FROM ONTARIO PUBLIC HEALTH UNITS

Anindya Sen

Anindya Sen*
Department of Economics
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1
asen@uwaterloo.ca

Soumis : 2021-02-05
Accepté : 2021-09-06

Abstract : From September 2020 onwards, COVID-19 cases have rapidly increased across Canada. This study estimates the effects of Google population mobility indicators on daily COVID-19 cases to evaluate the impacts of public movements across different regions in Ontario. We focus on Ontario as Google mobility data are available for Public Health Units (PHUs) for that province. Results based on pooled data from May 1st – November 15th imply that higher mobility at retail stores is significantly correlated with an increase in daily COVID-19 cases. However, empirical estimates from individual PHU level time-series models reveal regional differences, as these findings are based primarily on the relationship between retail mobility and daily cases for the Public Health Units (PHUs) of Toronto and Peel. These results support the implementation of region-specific lockdowns. Further, different specifications generate daily COVID-19 forecasts for Peel and Toronto that are on average, approximately 6%-9% different from actual values. The models of this research should be of value to local health authorities who are in search of simple models that are not computationally intensive and are capable of generating reliable forecasts for specific regions.

Keywords : COVID-19; Population Mobility; Google Data; Econometric Models; Ontario; Public Health Units

* I would like to thank an anonymous referee for useful comments. All errors remain my responsibility.

INTRODUCTION

Ontario has experienced a sharp rising trend in daily COVID-19 cases from September 2020 onwards, far beyond levels seen during the early days of the pandemic.[1] This is the case for other provinces as well. In response, provincial governments re-imposed policies intended to restrict movement in public spaces and have enacted different levels of lockdowns. These measures are based on the belief that restricting social mobility will lead to enhanced physical distancing and therefore, fewer infections. Hence, evaluating the effects of public mobility on daily COVID-19 cases is of first order policy importance.

This paper employs Google population mobility indicators to capture regional changes in public movements and assess their statistical significance in explaining trends in daily cases across the largest 12 Public Health Units (PHUs) in Ontario (constituting roughly 85% of the population). Evaluating regional differences in the relationship between community mobility and daily cases assumes prominence, given that Ontario has implemented different levels of lockdown measures across regions – through a 'colour-designation' strategy - conditional on the severity of cases and stress on local public health resources.[2] This research contributes to the literature by using an econometric framework to identify which regions have been the most adversely impacted by increased population mobility through higher COVID-19 cases. Our endeavour has been facilitated by Google's release of daily social mobility data through its COVID-19 'Community Mobility Reports.' As detailed on its website, Google creates these aggregated and anonymized sets of data from users who have turned on the Location History setting of Google accounts on their phones and have agreed to share this information. Using these data have the potential to shed light on the precise impacts of mobility on the spread of infections across different regions. Further, given that Google indicators are specific to mobility at different destinations such as retail stores and groceries, the results of this study may give some insight on the type of mobility that is most correlated with increases in daily cases. This is important, as almost one year into the pandemic, there is limited knowledge on what type of population movements are conducive to the spread of infections.

The motivation for focusing on regions in Ontario goes beyond the fact that the province has the highest number of cases in Canada. First, Ontario is also the only province which releases data on daily cases by test date. Having data by test date is critical to ensure a better matching between social mobility indicators and daily cases, as relying on case count by test release date would lead to confounded estimates if there is a significant time lag between tests and the release of results. Second, Google has released social mobility indicators for health regions in Ontario, but not for most other Canadian provinces. Hence, an attempt to identify the regional effects of population mobility on daily COVID-19 cases in Canada with Google indicators, is best accomplished with Ontario data.

There are quite a few studies that have attempted to forecast long-term case trajectories in Canada and across the world.[3] In contrast to models of long-term forecasts, there is much less work on short-term predictions.[4] Further, I am only aware of one other paper that investigates the effects of population mobility on COVID-19 cases in Canada. Specifically, Karaivanov et al. (2020) pool data across Ontario health regions and Canadian provinces to evaluate the effects of mask mandates and other non-pharmaceutical interventions (NPI) on COVID-19 case growth. While the paper relies on Google mobility data to account for behavioural decisions in mobility, it does not use the data to forecast daily COVID-19 cases. Some recent U.S. based studies investigate the effects of public policies on social mobility.[5] One paper that does estimate the relationship between population mobility and daily cases is Glaeser at al. (2020), who employ Safegraph data for some U.S. cities, but do not offer short-term predictions.[6] This research estimates the effects of population mobility indicators on daily cases across Ontario PHUs, and also assesses whether such models can be employed to produce accurate predictions.

DATA

Daily COVID-19 Cases

In Canada, Public Health Units (PHUs) are administrative areas consisting of cities and adjoining suburbs that are charged with overseeing and managing public health according to policies and directives issued by provincial ministries of health. Being the largest province in terms of population, Ontario has the most health regions (36). This paper employs data for the following twelve public health units (with population in parentheses): Durham (645,862); Hamilton (1,399,073); Halton (548,430); Middlesex-London (455,526); Ottawa (1,306,249); Niagara (447,888); Peel (1,381,744); Simcoe-Muskoka (540,249); Waterloo (535,154); Windsor (398,953); Toronto (2,731,571); and York (1,109,909). Together, these health units account for more than 85% of the province's population.[7] Toronto, Ottawa, Peel, Durham, and York also happen to be among Canada's top ten Census Divisions. Other PHUs are much smaller in population and did not experience a significant number of COVID-19 cases.

Figure 1 captures daily COVID-19 trends over the sample period (May 1st – December 15th) for the PHUs with the largest populations (Hamilton, Toronto, Peel, York, Ottawa and Durham) while Figure 2 contains trends in other smaller PHUs (Halton, Waterloo, Middlesex-London, Niagara, Simcoe, and Windsor).[8] Toronto and Peel have consistently had the highest daily cases, with a sharp increasing trend from around September 17th. York and Durham are next in order of the magnitude of daily cases. Relative to its population size, recent daily cases in Ottawa have been remarkably limited, with daily cases during the first two weeks of December ranging from 35 to 52. With respect to smaller PHUs, Windsor had a one-day spike on June 24th because of an outbreak among agricultural workers.[9] Aside from this spike, corresponding trends in other PHUs in Figure 2 were relatively stable till October, but then began to increase rapidly through November. Halton, Windsor, and Waterloo are the regions with the highest number of recent cases.

1 Please see Flanagan (2020) for further details.

2 Ontario currently uses the following colour codes in designating COVID-19 lockdown measures for PHUs: Green (Prevent) ; Yellow (Protect); Orange (Restrict); Red (Control); and Gray (Lockdown). Refer to 'COVID-19 response framework: keeping Ontario safe and open' available at https://www.ontario.ca/page/covid-19-response-framework-keeping-ontario-safe-and-open (last accessed May 9th 2021) for further information.

3 Ogden et al. (2020) and Tuite et al. (2020) are examples of Canadian research.

4 Altieri et al. (2020) use U.S. county level data to construct different types of linear and exponential predictors that are able to forecasts of COVID-19 deaths. Liu et al. (2020) use Bayesian methods to estimate a reduced form panel data model to generate cross-country forecasts of COVID-19 infections.

5 For example, see Barrios et al. (2020) and Goolsbee and Syverson (2020). With respect to Canada, Chan (2020a, 2020b) and Armstrong II et al. (2020) employ mobility data but do not predict trends in daily COVID-19 cases.

6 Alternatively, Mavragani & Gkillas (2020) use Google Trends data to predict U.S. COVID-19 cases.

7 Population data are downloadable from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

8 Data available from https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool.

9 Please see https://www.ctvnews.ca/health/coronavirus/advocates-demand-ontario-shut-down-farms-as-covid-19-cases-soar-among-workers-1.5004897 for further details (last accessed January 20th 2021).

Figure 1. Daily COVID-19 Cases Larger PHUs

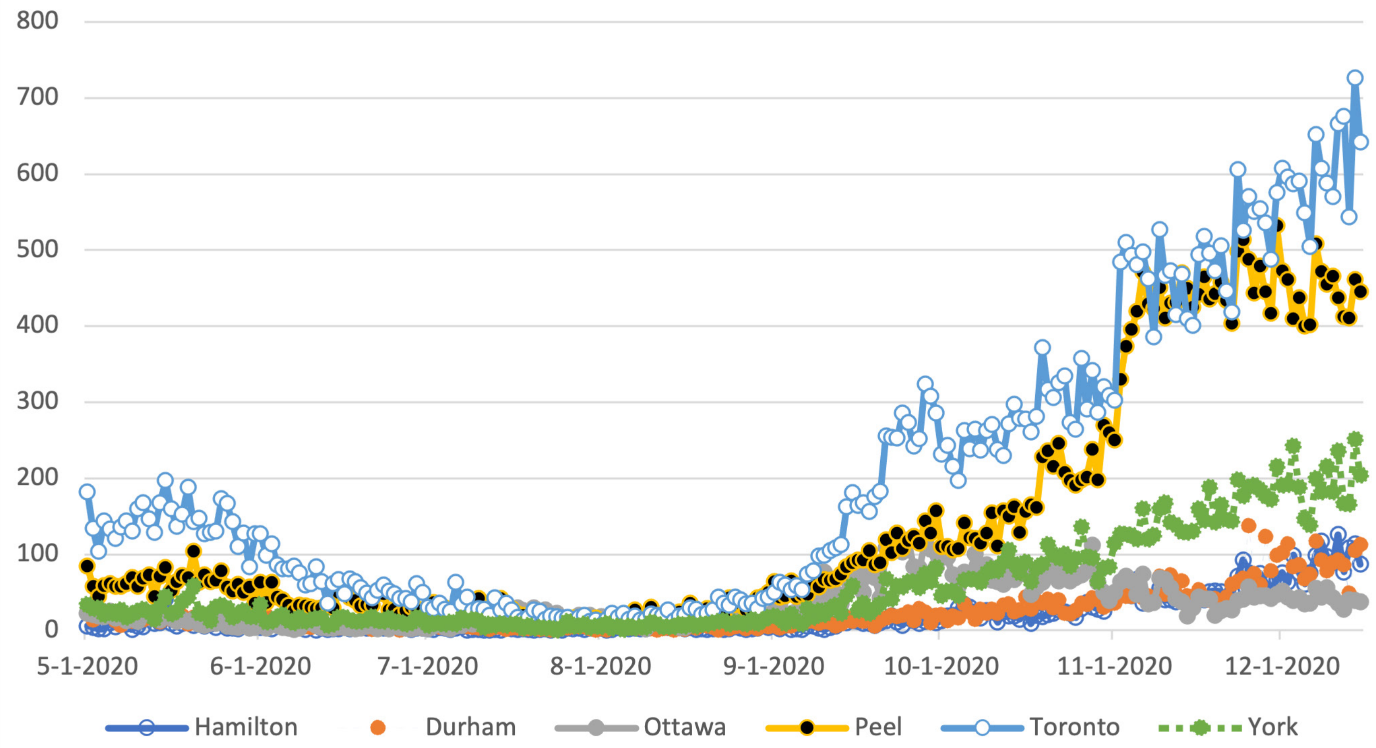
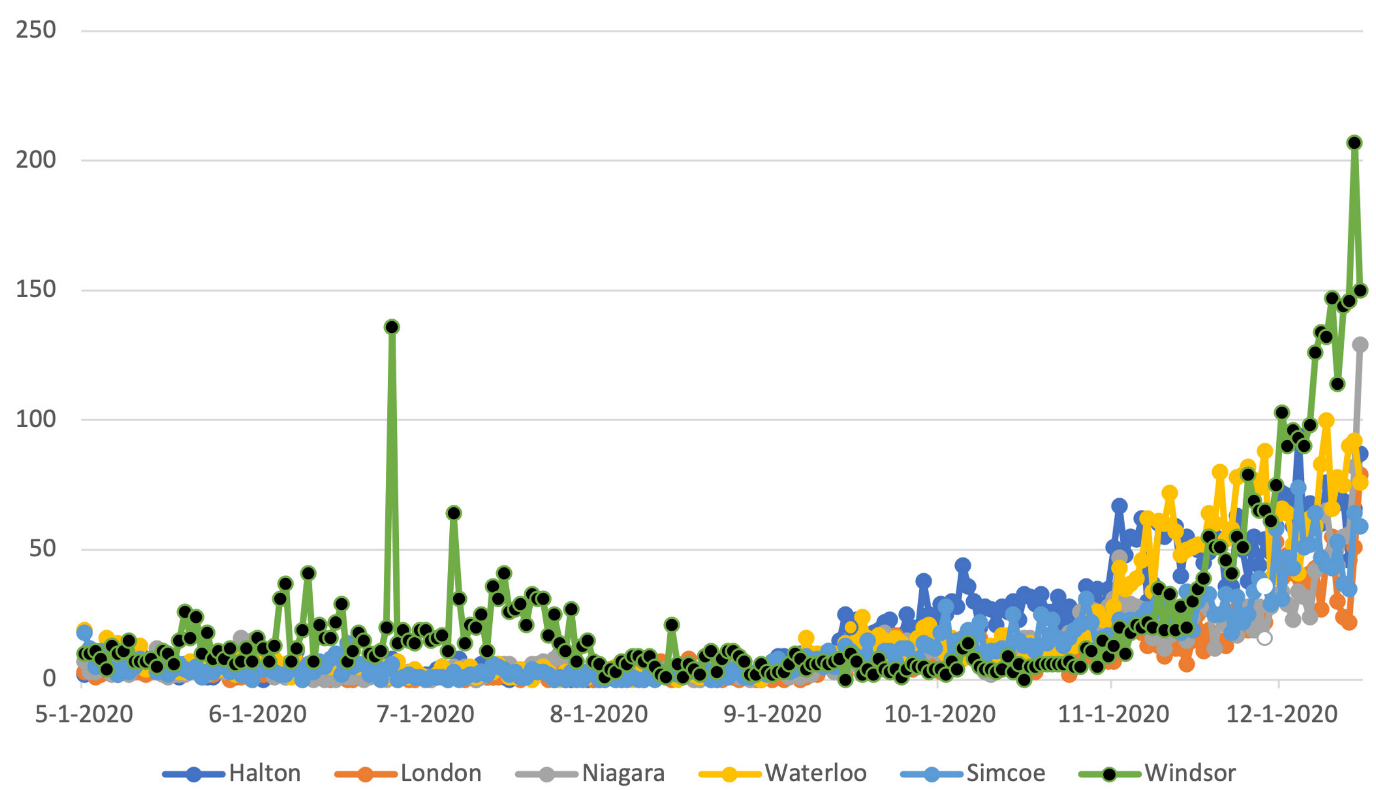


Figure 2. Daily COVID-19 Cases Smaller PHUs



Notes: Google mobility data were downloaded from https://www.google.com/covid19/mobility/

Figure 3. Google Retail Mobility at Larger PHUs

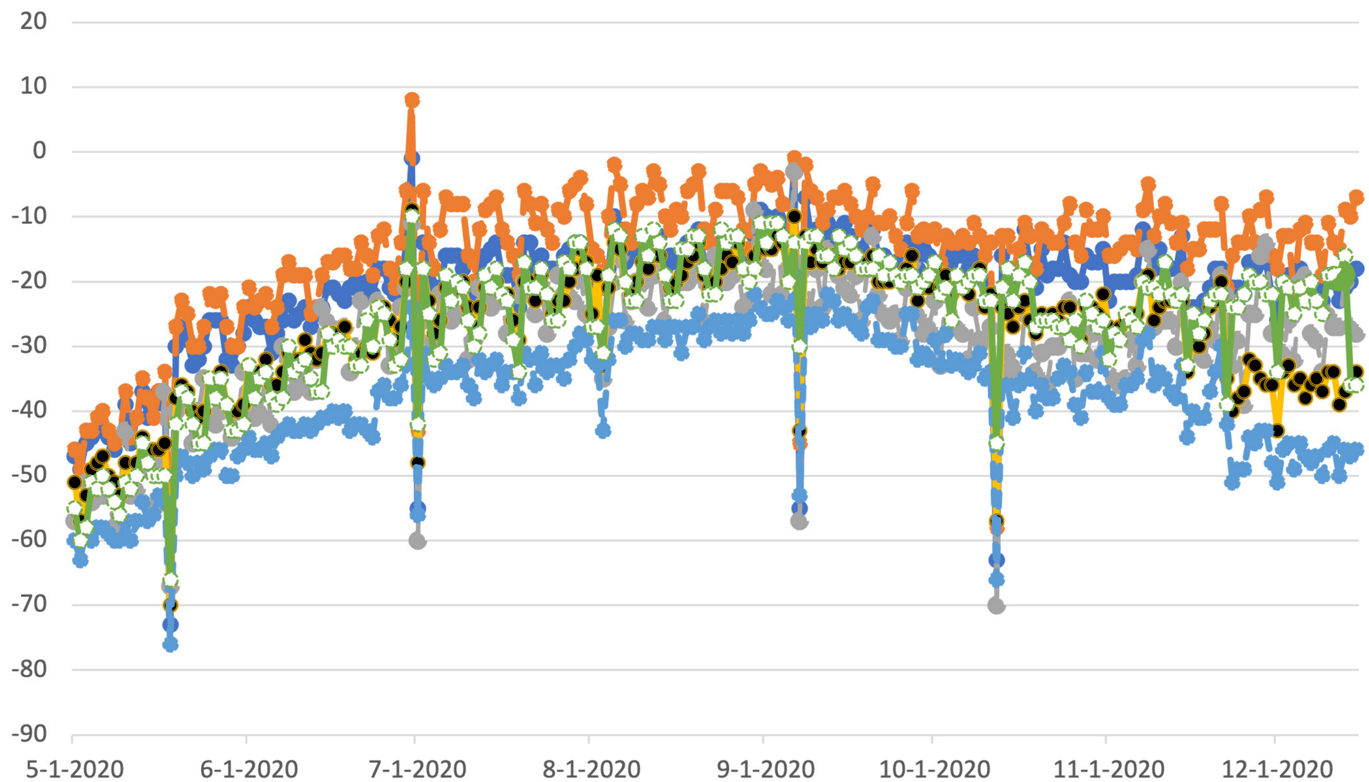
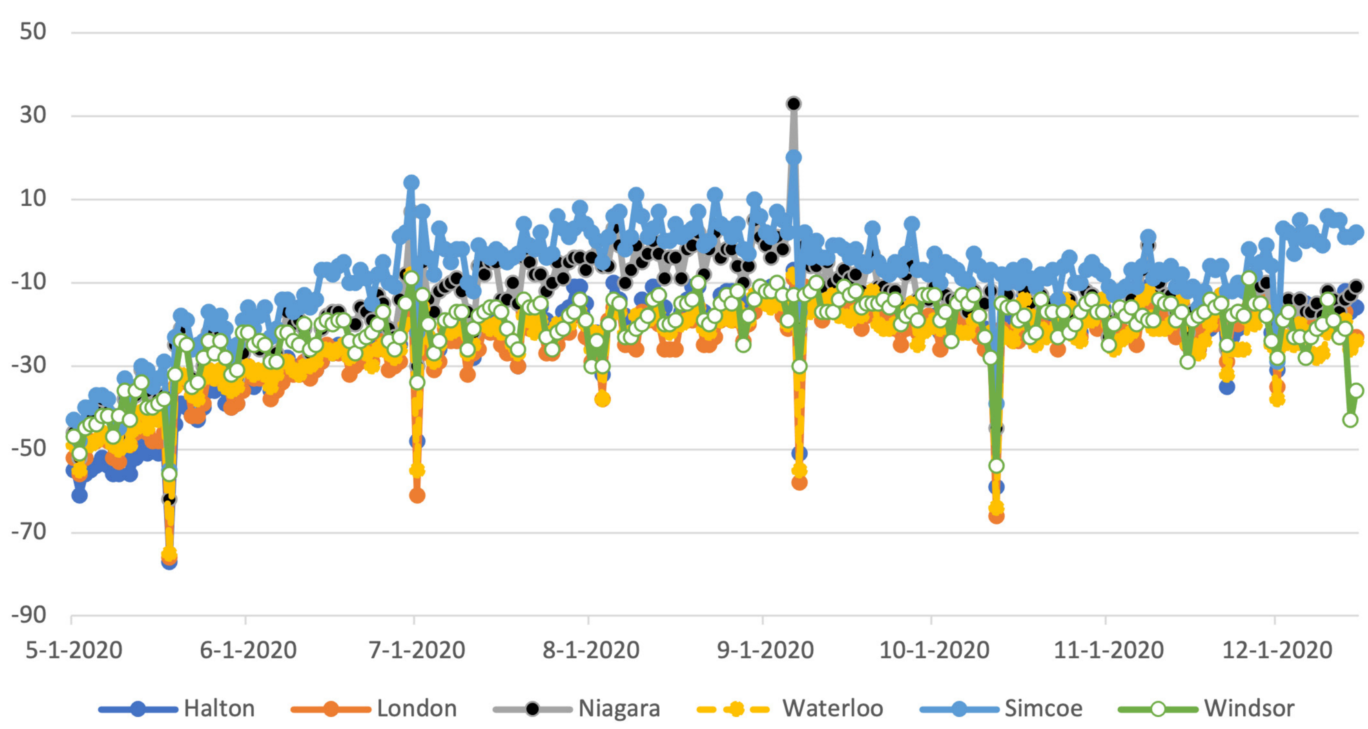


Figure 4. Google Retail Mobility at Smaller PHUs



Notes: Google mobility data were downloaded from https://www.google.com/covid19/mobility/

Figure 5. Google Grocery Mobility for Larger PHUs

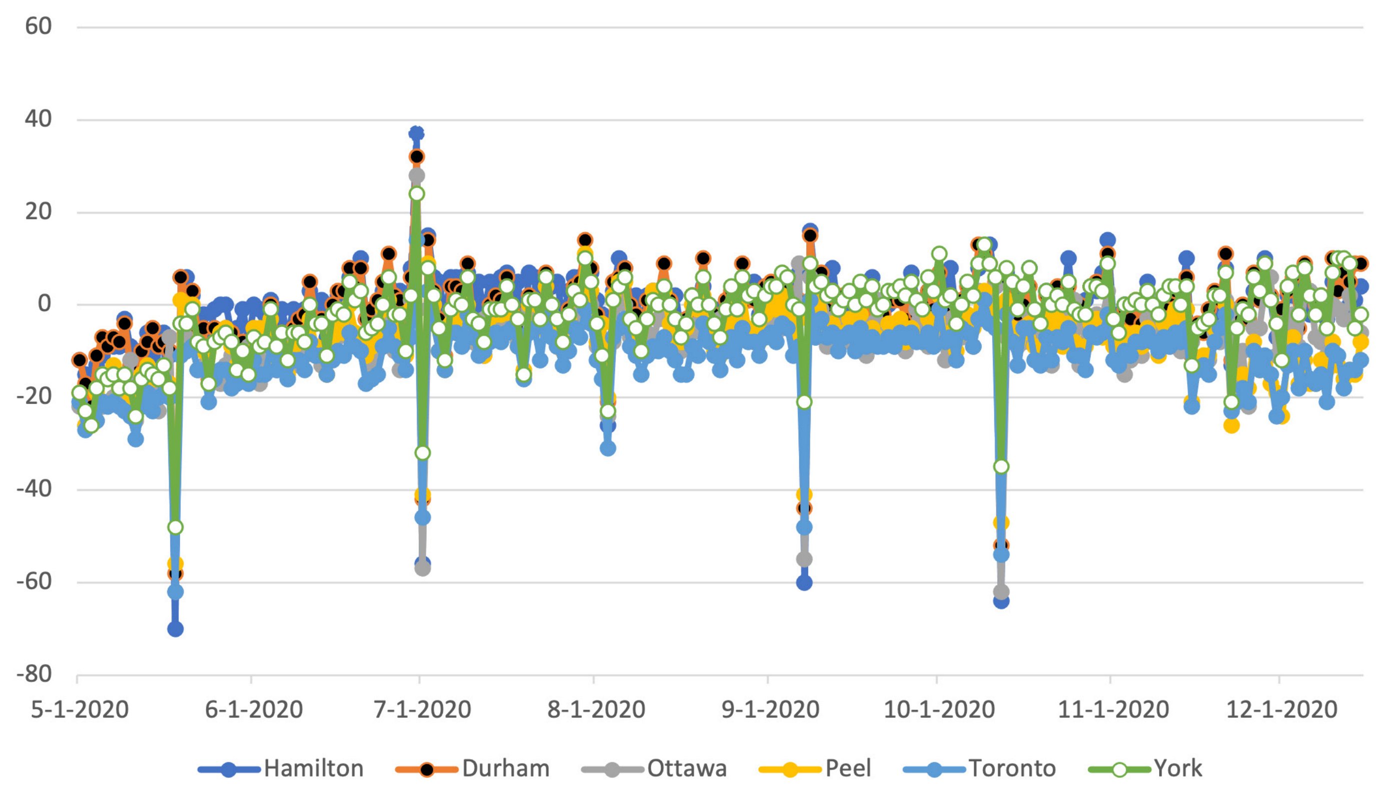
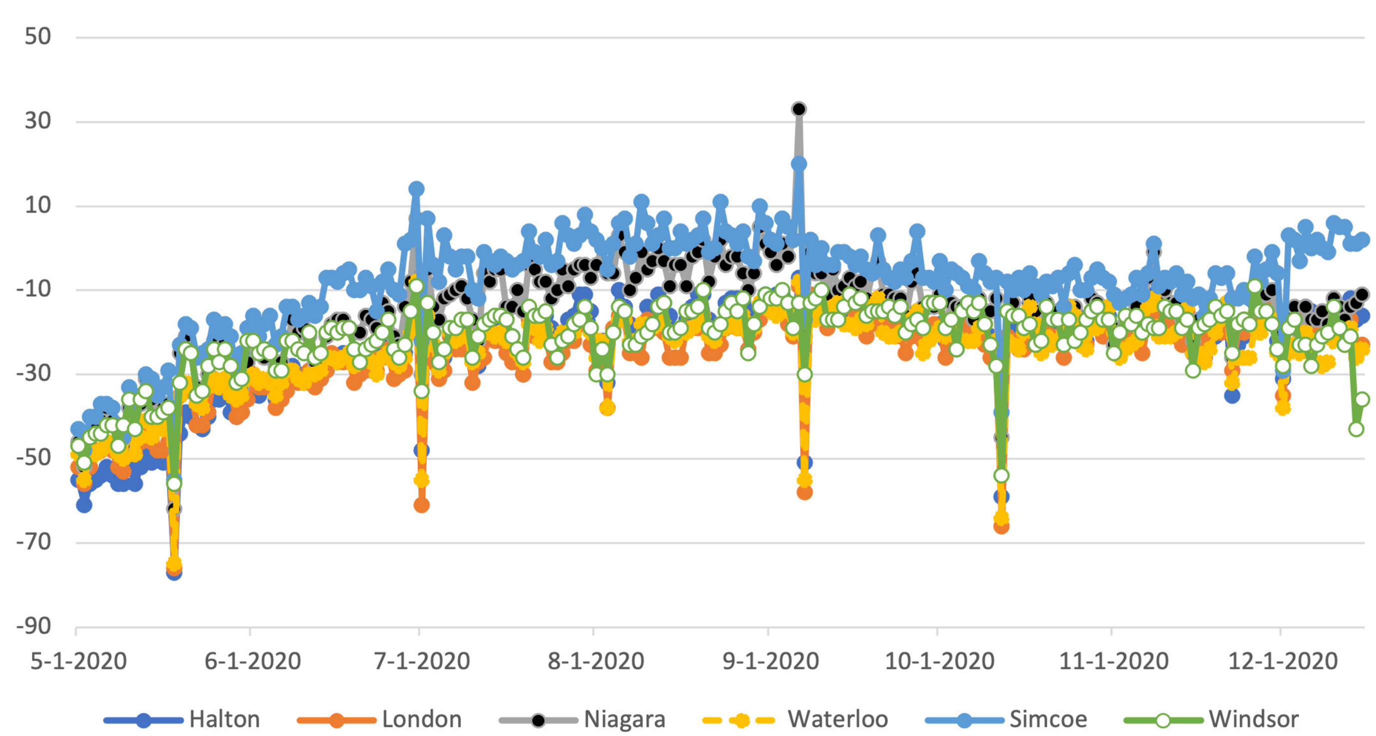


Figure 6. Google Grocery Mobility at Smaller PHUs



Notes: Google mobility data were downloaded from https://www.google.com/covid19/mobility/

Table 1. Summary Statistics May 1st – Nov 15th 2020 (2,388 observations)

Name	Mean	St. Dev	Variance	Min	Max
Cases by Episode Date	31.026	63.416	4021.6	0.0000	563.00
Seven Day Lagged Grocery Mobility	-3.5992	11.452	131.16	-73.000	48.000
Seven Day Lagged Retail Mobility	-24.487	14.398	207.30	-77.000	33.000
Business Closures/Openings Index	0.38655	0.28073	0.0788	0.18750	0.99375
Masks Dummy	0.6298	0.48296	0.2333	0.0000	1.0000
Weekend	0.31658	0.46524	0.21645	0.0000	1.0000

Google Population Mobility Indicators

Google has released population mobility indicators that capture total visits to destinations commonly frequented by individuals (https://www.google.com/covid19/mobility/). All these indicators are in the form of indices with values calculated relative to the baseline, which is the median for the corresponding day of the week, during the 5-week period Jan 3–Feb 6, 2020. These indices are daily values that are aggregated across individuals who have enabled their location history to be accessed by Google. Mobility indicators are employed for visits to retail and recreation venues (restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters) and grocery and pharmacy stores (grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies). Google also has indicators for visits to: (1) residences; (2) workplaces; (3) parks; and (4) transit hubs. We could not use data on parks and transit hub mobility due to missing values for multiple regions. With respect to residences and workplaces, coefficient estimates of these variables with respect to daily COVID-19 cases were not statistically significant in LASSO regressions, and hence not employed in the econometric analysis of this study.

Figure 3 (4) displays trends in Google retail mobility for the larger (smaller) PHUs in the sample. In general, retail mobility began to rise steeply until around early July, with more gradual increases up to early September. For the remainder of the sample period including October, retail mobility began to drop. Among large PHUs, retail mobility is lower in the high cases regions of Toronto and Peel. Similarly, high cases PHUs in smaller areas such as Halton and Waterloo have lower amounts of retail mobility than other comparable regions. Figures 5 and 6 captures movements in grocery mobility. In contrast to retail mobility, grocery mobility remained relatively stable over the sample period, with steep drops during long weekends and holidays. As is the case with retail mobility, grocery mobility is lower in Peel and Toronto in comparison to other large PHUs, and in fact started dropping during November. Similarly, high case small PHUs such as Halton and Waterloo also had lower grocery mobility than other comparable PHUs.

ECONOMETRIC MODEL

To further investigate the relationship between the number of Covid cases and the mobility data, the following empirical model is adopted;

$$\text{DailyCovidCases}_t = \alpha + \Sigma_k\gamma_k\text{DailyCovidCases}_{t-k} + \Sigma_k\theta_k\text{Retail}_{t-k} + \Sigma_k\upsilon_k\text{Grocery}_{t-k} + W_t + \varepsilon_t \tag{1}$$

The daily number of new confirmed cases on day 't' (DailyCovidCases_t) based on date of specimen collection is modeled as a function: of lagged values of itself ($\text{DailyCovidCases}_{t-k}$); lagged values of retail and recreation mobility (Retail_{t-k}); grocery mobility (Grocery_{t-k}); and a dummy variable representing weekends and holidays (W_t). γ_k, θ_k, and υ_k are the coefficients of these covariates, and ε_t is the error term, which is assumed to be identically and independently distributed. Table 1 contains summary statistics of all variables.

Equation (1) is first estimated globally with observations from all PHUs pooled together and then separately for each PHU in order to understand differences in the effects of social mobility on daily cases across jurisdictions. Given the time-series nature of the data and the likelihood of serially correlated errors, equation (1) is estimated using a Cochrane-Orcutt estimator. In all cases, the null hypothesis of no first order serial correlation could not be rejected on the basis of Durbin Watson tests. The model is trained using data from May 1st – November 15th 2020 and the robustness of specifications is evaluated through predictions based on data from November 16th – December 15th 2020.

A key challenge is to understand the appropriate lag structure for the lagged dependent variable and each social mobility indicator. According to WHO guidelines, the time from exposure to COVID-19 to the emergence of symptoms can begin, on average, 5-6 days, but range from 1-14 days (https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19.).[10] A data-driven strategy based on these guidelines and statistical significance, was used to choose which Google social mobility indicators to employ, along with corresponding lag structure. First, the statistical significance of six to twelve day lagged values of all social mobility indicators and one to four day lags in the dependent variable, were investigated in separate regressions for each PHU. Coefficient estimates of six day and ten to twelve day lagged values in social mobility indicators and three to four lags in daily cases, were consistently insignificant and hence dropped. Second, model fit was further investigated by employing LASSO regressions that included month dummies as well, to arrive at parsimonious specifications and avoid overfitting. One and two day lags of dependent variables remain statistically significant as well as seven day lags of retail and grocery mobility. In this respect, an obvious omitted right hand side variable is the number of daily tests. Unfortunately, testing data are unavailable at the PHU level. However, as a sensitivity test, a similar regression was conducted at the province level. Daily testing numbers were statistically insignificant on the inclusion of one and two day lags of daily COVID-19 cases as independent variables.

RESULTS

Baseline Estimates

Table 2 contains WLS estimates of the relationship between daily COVID-19 cases and Google mobility indicators obtained by pooling data across PHUs and over time, which yields on-average es-

10 Glaeser at al. (2020) use a 14 day mobility lag.

Table 2. Weighted Least Squares (WLS) Estimates of the Effects of Non-Pharmaceutical Interventions (NPIs) and Google Mobility on Daily COVID-19 Cases

	(1)	(2)	(3)
One Day Lagged Cases		0.655 (0.019)[a]	0.646 (0.0195)[a]
Two Day Lagged Cases		0.348 (0.019)[a]	0.341 (0.02)[a]
Business Closures/Openings Index	-190.98 (13.53)[a]	-3.306 (2.956)	-0.017 (3.683)
Mask Mandate	44.164 (6.715)[a]	0.129 (1.421)	-1.998 (1.728)
Seven Day Lagged Retail Mobility	-7.089 (0.287)[a]	0.136 (0.068)[b]	0.194 (0.091)[b]
Seven Day Lagged Grocery Mobility	3.22 (0.294)[a]	-0.224 (0.064)[a]	-0.247 (0.072)[a]
City Fixed Effects	No	No	Yes
Weekend/Holiday Fixed Effect	Yes	Yes	Yes
Quarterly Fixed Effects	No	No	Yes
Adjusted R Square	0.2920	0.9689	0.9690
Observations	2,388	2,388	2,388

Notes: The regressions in this table are based on data pooled across 12 Public Health Units (PHUs) in Ontario from May 1st- November 15th 2020. They are: Hamilton, Durham, Ottawa, Peel, Toronto, York, Halton, Middlesex-London, Niagara, Simcoe-Muskoka, Windsor, and Waterloo. Standard errors are in parentheses below coefficient estimates and are robust and Newey-West corrected for autocorrelation. a, b, and c denote statistical significance at the 1%, 5%, and 10% levels.

timates of the effects of retail and grocery mobility that are conditioned on the use of PHU specific fixed effects.[11] Besides providing a baseline, this exercise allows an assessment of whether time-series regressions run individually for each PHU and based on equation (1), might result in confounded estimates due to unobserved heterogeneity. Running regressions based on pooled data also allows an assessment of the impacts of non-pharmaceutical interventions (NPIs) such as mandatory mask use in indoor spaces, and business closures and restrictions on social gatherings. There is time-series variation across PHUs in the implementation of these initiatives, which enables an evaluation of their efficacy. Mandatory mask usage is measured through a dummy variable while business closures are captured through an index variable bounded by 0 and 1, with values close to 1 denoting widespread closure of non-essential businesses and restrictions on social gatherings.[12] Weighted Least Squares (WLS) regressions are run with observations weighted by PHU specific population to control for the obvious effects of larger population on the incidence of daily cases. Further, controlling for the implementation of compulsory indoor mask usage and restrictions on business operations and social gatherings, serves as a potential check for endogeneity bias in coefficient estimates of population mobility, as NPIs are a response to rising daily cases.[13]

Column (1) focuses on the effects of NPIs and Google mobility indicators, column (2) evaluates the impacts of including lagged dependent variables, and column (3) includes city and quarterly specific dummies. In column (1), indoor mask usage is positively and significantly correlated (at the 1% level) with daily cases, while the business and gatherings restrictions index is negative and statistically significant (at the 1% level). Further, coefficient estimates of retail and grocery mobility are also statistically significant at the 1% level. However, both indoor mask usage and business/gatherings becomes insignificant in remaining columns with the addition of other covariates. In columns (2) and (3), lagged cases are positive and statistically significant at the 1% level. Estimates in column (3) imply that a 10-percentage point increase in retail social mobility is associated with a roughly 1.9 case rise on a per day basis, while a 10-percentage point increase in grocery mobility is correlated with roughly a 2.5 case decline on a daily basis.

Individual Time Series Regressions

The results of individual time-series regressions done with daily data for each PHU from May 1st – November 15th 2020 are contained in Table 3. The dependent variable in Table 3 is the number of daily cases by date of specimen collection. Panel A contains results from specifications for large PHUs, while corresponding results from smaller PHUs are detailed in Panel B. The first observation from Panel A is that our relatively simple specification fits the data quite well, with adjusted R squares ranging from 0.8 to 0.98. Second, coefficient estimates of the effects of lagged cases are consistent with findings in Table 2, as an increase in one and two day lagged cases is significantly correlated with higher current cases , at the 1% - 10% levels. Third, coefficient estimates of retail mobility are positive and statistically significant at the 5% levels of significance for Peel, Toronto, and York, and at the 10% level for Ottawa. A 10 percentage point rise in retail mobility is associated with roughly a 2, 3, 4, and 8 case per day increase in Ottawa, York, Peel, and Toronto, respectively. Fourth, an increase in grocery mobility is significantly correlated at either the 1% or 5% levels with lower daily cases for most of these cities.

Except for Middlesex-London and Windsor, results from Panel B reveal high model fits for most smaller PHUs with adjusted R squares from 0.84 - 0.92. One day lagged cases are statistically significant at the 1% level for all PHUs. Coefficient estimates of two day lags are not significant for all PHUs and are generally smaller in magnitude relative to estimates in Panel A. Population mobility variables are significant at varying levels for Middlesex-London, Simcoe-Muskoka, and Waterloo. A 10 percentage point rise in retail mobility is associated with approximately an one case per day increase in these PHUs. Coefficient estimates of grocery mobility are negative and statistically significant for Middlesex-London and Waterloo.

11 Standard errors of these coefficient estimates are robust and Newey-West corrected for first order autocorrelation. These regressions also include quarterly dummies to account for unobserved seasonal effects.

12 Specific implementation dates of indoor mandatory mask usage were taken from Karaivanov et al. (2020), which is also the source for the business closures index. For the PHUs in this study, dates of mandatory mask usage are: Durham (July 10, 2020); Halton (July 22, 2020); Hamilton (July 20, 2020); Middlesex-London (July 18, 2020); Niagara (July 31, 2020); Ottawa (July 07, 2020); Peel (July 10, 2020); Region of Waterloo (July 13, 2020); Simcoe Muskoka (July 13, 2020); Toronto (July 07, 2020); Windsor-Essex (June 26, 2020); and York Region (July 17, 2020). The index variable captures regulations and restrictions on non-essential businesses and retail, personal businesses, restaurants, bars and nightclubs, places of worship, events, gyms and recreation, and limits on gatherings.

13 I am grateful to an anonymous referee for recommending the use of NPIs as a sensitivity check.

Table 3. Cochrane-Orcutt Estimates of the Effects of Google Mobility on Daily COVID-19 Cases

Large PHUs	Hamilton	Durham	Ottawa	Peel	Toronto	York
One Day Lagged Cases	0.670 (0.121)[a]	0.58 (0.107)[a]	0.595 (0.145)[a]	0.641 (0.206)[a]	0.604 (0.254)[a]	0.606 (0.149)[a]
Two Day Lagged Cases	0.282 (0.116)[b]	0.397 (0.104)[a]	0.331 (0.135)[b]	0.371 (0.206)[c]	0.392 (0.250)[c]	0.416 (0.149)[a]
Seven Day Lagged Retail Mobility	0.041 (0.052)	0.065 (0.043)	0.243 (0.126)[c]	0.418 (0.197)[b]	0.776 (0.339)[b]	0.284 0.129)[b]
Seven Day Lagged Grocery Mobility	-0.0498 (0.061)	-0.064 (0.061)	-0.182 (0.157)	-0.656 (0.267)[b]	-0.807 (0.484)[b]	-0.480 (0.178)[a]
Weekend Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R Square	0.8040	0.8423	0.8509	0.9803	0.9775	0.9441
Root Mean Square Error (RMSE)	20.97	24.63	9.39	41.092	61.66	36.58
Mean Absolute Percentage Deviation (MAPD)	21.92%	22.61%	20.33%	7.11%	6.24%	16.79%

Smaller PHUs	Halton	Middlesex-London	Niagara	Simcoe-Muskoka	Windsor	Waterloo
One Day Lagged Cases	0.747 (0.110)[a]	0.815 (0.210)	0.940 (0.227)[a]	0.463 (0.091)[a]	0.429 (0.167)[b]	0.801 (0.099)[a]
Two Day Lagged Cases	0.244 (0.11)[b]	0.115 (0.169)	0.163 (0.167)	0.500 (0.087)[a]	0.147 (0.101)	0.199 (0.102)[b]
Seven Day Lagged Retail Mobility	0.031 (0.048)	0.0899 (0.032)[b]	0.038 (0.03)	0.0913 (0.052)[c]	-0.094 (0.120)	0.134 (0.048)[b]
Seven Day Lagged Grocery Mobility	-0.008 (0.071)	-0.118 (0.038)[a]	-0.048 (0.042)	-0.107 (0.079)	0.126 (0.202)	-0.1521 (0.058)[b]
Weekend Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R Square	0.8972	0.5536	0.8384	0.8470	0.1563	0.9215
Root Mean Square Error (RMSE)	14.03	11.34	14.88	11.66	24.72	15.10
Mean Absolute Percentage Deviation (MAPD)	18.51%	25.38%	28.57%	22.89%	25.68%	18.26%

Notes: The regressions in this table are based on data from May 1st- November 15th 2020 consisting of 229 observations for each Public Heath Unit. Regression estimates are obtained from Cochrane-Orcutt transformations. Standard errors are in parentheses below coefficient estimates. a, b, and c denote statistical significance at the 1%, 5%, and 10% levels. Root Mean Square Errors (RMSE) and Mean Absolute Percentage Deviations (MAPDs) are constructed with respect to data from November 16th – December 15th.

Table 4. Cochrane-Orcutt Estimates of the Effects of Google Mobility on Natural Logarithm of Daily COVID-19 Cases

	Hamilton	Durham	Ottawa	Peel	Toronto	York
One Day Lagged Cases	0.423 (0.107)[a]	0.421 (0.141)[a]	0.609 (0.130)[a]	0.588 (0.097)[a]	0.702 (0.133)[a]	0.537 (0.124)[a]
Two Day Lagged Cases	0.199 (0.080)[b]	0.103 (0.087)	0.345 (0.125)[a]	0.408 (0.096)[a]	0.294 (0.133)[b]	0.1791 (0.095)[c]
Seven Day Lagged Retail Mobility	0.046 (0.036)	-0.017 (0.033)	0.006 (0.005)	0.005 (0.002)[b]	0.008 (0.002)[a]	0.002 (0.016)
Seven Day Lagged Grocery Mobility	-0.099 (0.043)[b]	-0.011 (0.045)	-0.002 (0.006)	-0.006 (0.003)[b]	-0.0102 (0.003)[a]	-0.0098 (0.023)
Weekend Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R Square	0.1549	0.0797	0.8899	0.9343	0.9797	0.2665
Root Mean Square Error (RMSE)	21.70	65.05	8.96	38.13	67.84	99.77
Mean Absolute Percentage Deviation (MAPD)	21.92%	70.34%	18.61%	6.37%	8.89%	48.81%

Notes: The regressions in this table are based on data from May 1st- November 15th 2020 consisting of 229 observations for each Public Heath Unit. Regression estimates are obtained from Cochrane-Orcutt transformations. Standard errors are in parentheses below coefficient estimates. a, b, and c denote statistical significance at the 1%, 5%, and 10% levels. Root Mean Square Errors (RMSE) and Mean Absolute Percentage Deviations (MAPDs) are constructed with respect to data from November 16th – December 15th

In summary, these findings reveal retail mobility to be a statistically significant determinant of COVID-19 incidence in four of the large PHUs (Toronto, Ottawa, Peel, and York), which have experienced particularly sharp increases in daily cases in recent months and three smaller regions (Middlesex-London, Simcoe-Muskoka, and Waterloo). The negative sign of grocery mobility might reflect endogeneity bias as increases in daily cases drive down visits to grocery stores. This is consistent with findings by Chan (2020b).

The specifications in Table 3 were also employed to construct predictions on daily cases from November 16th – December 15th. The ability to generate accurate forecasts is important, as it yields an idea on future trends for resource planning purposes. Forecasts were constructed with data updating on a weekly basis, and the corresponding Root Mean Square Errors (RMSEs) and the Mean Absolute Percentage Deviations (MAPDs) are reported for each Public Health Unit.[14] Larger PHUs for which Google retail indicators are statistically significant – Toronto, Peel, and York - have lower MAPDs relative to other PHUs, ranging from roughly 6% - 17%. As a sensitivity analysis, Table 4 contains alternative estimates with the natural logarithm of daily cases as the dependent variable. This was done to evaluate the existence of a non-linear relationship between daily cases and population mobility. Only the results with respect to large PHUs are reported as the adjusted R squares for smaller PHUs were extremely low. In terms of statistical significance, one and two day lagged cases are statistically significant for most cities and retail and grocery mobility are significant for Toronto and Peel. MAPDs are comparable to corresponding values for Toronto, Peel, Ottawa, and Hamilton, but much higher for Durham and York.

14 The MAPD is calculated by taking the percentage difference of each daily forecast from its actual value between November 16th-December 15th, converting it to an absolute value and then calculating the mean across this sample range.

CONCLUSION

The use of Google mobility indicators enables an assessment of the types of population movements that have contributed towards rising daily cases. Empirical estimates for individual PHUs imply significant regional differences in the effects of population mobility. First, higher retail mobility is correlated with an increase in daily cases for the largest PHUs of Toronto and Peel, which have also been experiencing the most significant increase in daily cases. This finding is robust to the use of different specifications. Results from levels models imply that increases in retail mobility are also associated with higher daily cases in Middlesex-London, Simcoe-Muskoka, and Waterloo. However, the results for smaller PHUs become insignificant when the dependent variable is in natural logarithms. Given the statistical insignificance of retail mobility in other regions, these findings support the implementation of retail closures on a regional basis to curb rising COVID-19 cases. Second, results from forecasting exercises imply that the larger PHUs in which retail mobility is statistically significant, also have lower Mean Absolute Percentage Deviations (MAPDs).

Our results should be of value to policymakers, as they yield models that are region specific, and can be used for resource allocation issues, such as preparing hospitals for an influx of patients. Further, they yield support for the extended closure of retail businesses in Toronto and Peel regions, given the robustness of coefficient estimates of retail mobility across different specifications. The models are simple enough that they can be easily built upon and appropriately modified by local policymakers with limited resources. The findings of this study demonstrate that Google data can be used to construct evaluate the effects of some types of mobility and construct reasonable predictions of day ahead COVID-19 cases.

REFERENCES

Altieri, N., Barter, R.L., Duncan, J., Dwivedi, R., Kumbier, K., Li, X., Netzorg, R., Park, B., Singh, C., Tan, Y.S., Tang, T., Wang, Y., Zhang, C. & Bin, Y. (2021). Curating a COVID-19 Data Repository and Forecasting County-Level Death Counts in the United States. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.1d4e0dae

Armstrong II, D., Lebo, M.J. & Lucas, J. (2020). Do COVID-19 Policies Affect Mobility Behaviour? Evidence from 75 Canadian and American Cities. *Canadian Public Policy* 46(S2): S127-S144. https://doi.org/10.3138/cpp.2020-062.

Barrios, J.M., Benmelech, E., Hochberg, Y.V., Sapienza, P. & Zingales, L. (2020). Civic Capital and Social Distancing During the Covid-19 Pandemic. *Journal of Public Economics* 193, https://doi.org/10.1016/j.jpubeco.2020.104310.

Chan, J. (2020a). The Geography of Social Distancing in Canada: Evidence from Facebook. *Canadian Public Policy* 46(S1): S19-S28.

Chan, J. (2020b). Using Google Data to Understand Canadian Movement Reductions During the COVID-19 Pandemic (May 12, 2020). Available at SSRN: https://ssrn.com/abstract=3599227 or http://dx.doi.org/10.2139/ssrn.3599227

Flanagan, R. (2020). Canada Has Twice as Many Active COVID-19 Cases as it Did on Nov. 1. Available at https://www.ctvnews.ca/health/coronavirus/canada-has-twice-as-many-active-covid-19-cases-as-it-did-on-nov-1-1.5209320, last accessed on January 15th 2021.

Glaeser, E.L., Gorback, C. & Redding, S.J. (2020). How Much Does COVID-19 Increase with Mobility? Evidence from New York and Four other U.S. cities. Forthcoming, *Journal of Urban Economics*.

Goolsbee, A. & Syverson, C. (2020). Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020. *Journal of Public Economics*, 193

Karaivanov, A., Shih E.L., Hitoshi, S., Chen, C. & Pamplona, S. (2021). Face Masks, Public Policies and Slowing the Spread of COVID-19: Evidence from Canada. *Journal of Health Economics* 78: 102475.

Liu, L.M., Hyungsik R. & Schorfheide, F. (2020). Panel Forecasts of Country-Level COVID-19 Infections., *Journal of Econometrics,* Available at https://laurayuliu.com/covid19-panel-forecast/ and forthcoming

Mavragani, A. & Gkillas, K. (2020). COVID-19 Predictability in the United States using Google Trends Time Series. *Scientific Reports* 10: 20693. https://doi.org/10.1038/s41598-020-77275-9

Ogden, N.H., Fazil, A., Arino, J., Berthiaume, P., Fisman, D.N., Greer, A.L., Ludwig, A., Ng, Victoria, T., Ashleigh R., Turgeon, P., Waddell, L.A. & Wu, J. (2020). Modelling Scenarios of the Epidemic of COVID-19 in Canada. *Artificial Intelligence in Public Health*, 46–6.

Tuite, A.R., Fisman D.N. & Greer, A.L. (2020). Mathematical Modelling of COVID-19 Transmission and Mitigation Strategies in the Population of Ontario, Canada. *Canadian Medical Association Journal (CMAJ),* 192 (19): E497-E505; DOI: https://doi.org/10.1503/cmaj.200476