The determinants of the differential exposure to COVID-19 in New York City and their evolution over time*

Milena Almagro† Angelo Orane-Hutchinson‡

April 21, 2020

Updated frequently. Click here for most updated version.

Abstract

In this paper we explore different channels to explain the disparities in COVID-19 incidence across New York City neighborhoods. To do so, we estimate several regression models to assess the statistical relevance of different variables such as neighborhood characteristics and occupations. Our results suggest that occupations are crucial for explaining the observed patterns, with those with a high degree of human interaction being more likely to be exposed to the virus. Moreover, after controlling for occupations, commuting patterns do not play a significant role. The relevance of occupations is robust to the inclusion of demographics, with some of them, such as income or the share of Asians, having no statistical significance. On the other hand, racial disparities still persist for Blacks and Hispanics compared to Whites albeit their effects are economically small. Additionally, we show that there is a selection effect with those residents in worse conditions being more likely to get tested. In our daily analysis, we show that this selection is considerably large at earlier dates but substantially decreases over time. While many occupations and demographics also become less important over time, we find effects consistent with higher intra-household contagion, in line with the progression of the pandemic as well as the health policies that have been set in place.

*Any errors or omissions are our own.

†Department of Economics, New York University. Email: m.almagro@nyu.edu
‡Department of Economics, New York University. Email: aoh227@nyu.edu

1 Introduction

There is extensive evidence of large differences on the incidence of COVID-19 across demographic groups and locations. For example, in the US African Americans are being hit the hardest by the pandemic, which creates an additional source of worry to this already vulnerable socio-economic group.[1] In the case of New York City (NYC), a report with race-specific data released on April 6 by the Department of Health and Mental Hygiene (DOH) revealed that Blacks and Hispanics have a 47% and a 36% higher crude rate, number of deaths per 100,000 population, compared to Whites respectively.[2] These patterns repeat themselves in other areas of the US, however the disparities across racial groups are less pronounced compared to NYC.[3]

The fact that more vulnerable groups and minorities are being hit harder by the pandemic has called the attention of many economists and policy makers. For example, Borjas (2020) and Schmitt-Grohé et al. (2020) show that incidence disparities of testing and positive rates across NYC neighborhoods is mainly explained by selection on demographics. On the other hand, as COVID-19 is a disease that intrinsically does not discriminate across demographics, the existing papers fail at pointing out the direct mechanisms that help explain such disparities. Hence, the goal of this paper is to shed some light on the importance of a set of such mechanisms beyond demographics, such as density, commuting patterns, and occupations.

To understand the relevance of different channels, we use data on the number of tests and positives across NYC zip codes provided by DOH.[4] We combine this data with demographic data provided by the American Community Survey (ACS), also at the zip code level.

Because our focus is to identify observable channels that are likely to explain why some demographic groups have a higher incidence to COVID-19 exposure, we estimate different specifications highlighting the importance of difference mechanisms at each stage. We start by including a small set of neighborhood controls, such as commuting patterns, population density, and health controls. In all of our specifications, we also include the number of tests per capita, to control for selection on testing. We find that when the number of tests per capita increases, the share of positives also increases. This result suggests that people in worse conditions are more likely to be tested, which is consistent with how testing has been handled in NYC, where only those who are admitted into a hospital have access to tests.[5] Surprisingly, we find that this

[1] For a recent article see `https://nyti.ms/2UUl5je`

[2] For the latest data release visit: `www1.nyc.gov/site/doh/covid/covid-19-data.page`

[3] For a summary see: `https://nyti.ms/2wu9jCK`

[4] Unfortunately, at the time of this analysis, there is no data available with the number of deaths by zip code.

[5] A selection from the patients can also lead to such pattern with those in worse conditions being more willing to get tested. However, given the way NYC has dealt with testing and the lack of empirical evidence supporting this argument, we believe that a selection on health providers on patients is more plausible reason for such results.

selection on testing becomes less important as days go by, which can be explained by the increasing availability of tests in NYC over time. Moreover, our analysis also shows that the importance of different channels change over time with many variables showing a clear temporal evolution, such as racial disparities for minorities becoming less important over time.

We continue by analyzing the role of occupations motivated by the fact that they vary in their degree of human interaction, with those with a high level of human contact being more likely to be exposed to the virus and vice versa.[6] We do so by including the share of workers in each zip code for thirteen categories constructed from ACS according to their degree of human interaction showing that indeed occupations are a key component in explaining the observed differences across NYC areas. For example, in our preferred specification including demographics and borough fixed effects, we find that increasing one percentage points the number of workers employed in transportation, an occupation that has been declared essential and with high exposure to human interaction, increases the share of positives by two and one percent for April 1 and April 20 respectively.[7] Moreover, we show that after controlling for occupations, length of commute and the use of public transport are not significant.[8]

Additionally, these results are robust to the inclusion of demographics as well as borough fixed effects.[9] Including demographics leads to several striking patterns. While simple correlations between income level and positive rates are negative, we show that income has no effect when occupations are included. However, we still see positive significant effects on positive rates for minorities, albeit barely economically relevant. Their magnitudes decrease over time as more testing becomes available, with Asians showing no statistical significance at the end of our sample. For April 1 and April 20, we find that increasing a percentage point in the population of Blacks correlates with an increase of 0.34% and 0.15% in the share of positives for an average number of 51% and 54% positive cases respectively. For Hispanics, the disparity is larger where a percentage point increase in their population corresponding to a 0.38% increase and a 0.23% in the rate of positives, respectively.

Our daily analysis also reveals that as the stay-at-home orders starts being effective, the effects of many occupations decrease their magnitude as days go by. However, we still a rather stable effect of household size over time consistent with the stay-at-home order being helpful at mitigating contagion at work or public spaces but not as much within the household.

We conclude that much of the disparities in the rates of positives can be explained by different demographic groups being more or less representative across different occupations. In particular, one key channel is the differences in exposures to human

[6]A recent paper by Barbieri et al. (2020) shows evidence of this mechanism for workers in Italy.

[7]The average rate of positives for those being tested was 51% and 54%.

[8]Harries (2020) argues that the NYC subway was crucial for spreading the pandemic in NYC. More recently, Furth (2020) shows that "local infections are negatively correlated with subway use."

[9]We use similar controls to those in Borjas (2020) for comparability purposes.

contact across jobs. An immediate implication for policy is to target first these more sensitive groups regarding the distribution of protective gear, testing, and vaccination should target first, not only considering their risk of exposure but also the potential spillovers that it may have on the rest of the population.

2 Data description and patterns

Our source of incidence rates of COVID-19 and the number of tests performed is the NYC Department of Health and Mental Hygiene (DOH) data release. The DOH releases the cumulative count of COVID-19 cases and the total number of residents that have been tested, divided by zip code of residence. This allows us to construct the tests per capita variable we use to control for selection in testing. Figure 1 gives an example of differences in rates of positives across NYC zip codes.

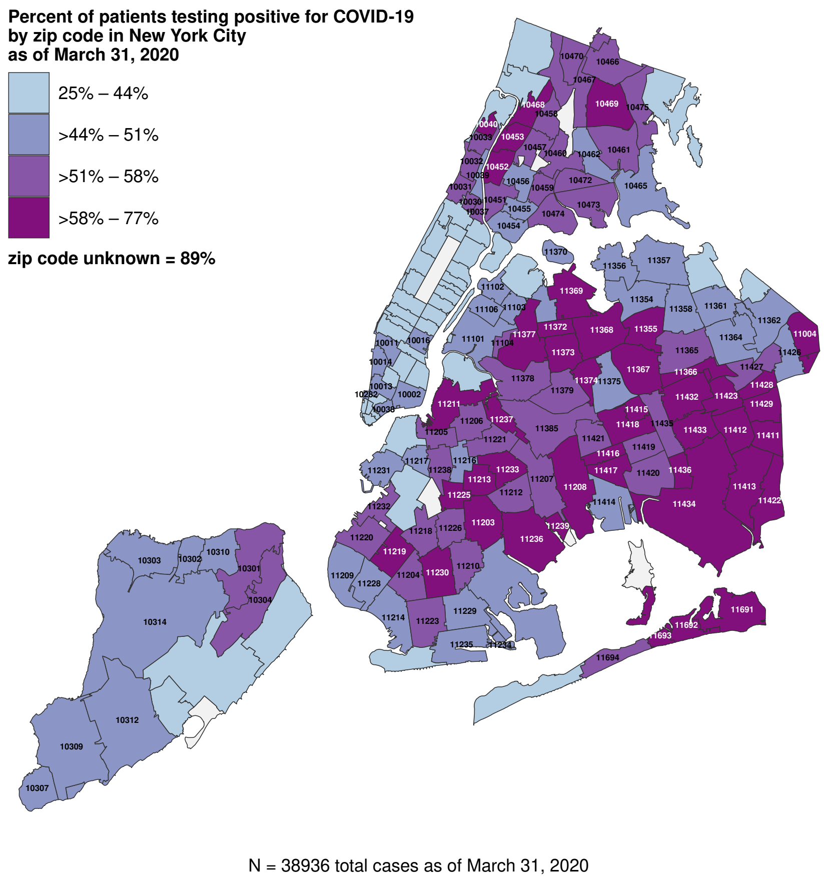

Figure 1: Map of rate of positives by zip code as of March 31, 2020.

From simple inspection, zip codes with the highest rates belong to the boroughs of Bronx, Brooklyn, and Queens. These boroughs are also home to the majority of

Blacks and Hispanics living in NYC.[10] We obtain demographic and occupation data at the zip code level from the American Community Survey (ACS). The demographic characteristics we include are zip code median income, average age, racial breakdown, and health insurance condition. We also include commuting related variables: average commuting time to work as well as means of transportation. Moreover, the zip codes with the lowest income levels are also located in those three boroughs We plot a simple correlation between the share of positives and demographics. We see that shares of Blacks and Hispanics are positively correlated with rate of positives, while a flat relationship for share of Asians as shown in Figure 2. We observe a negative relationship for income, slightly positive for share of males, and positive and significant for household size, as shown in Figure 3.

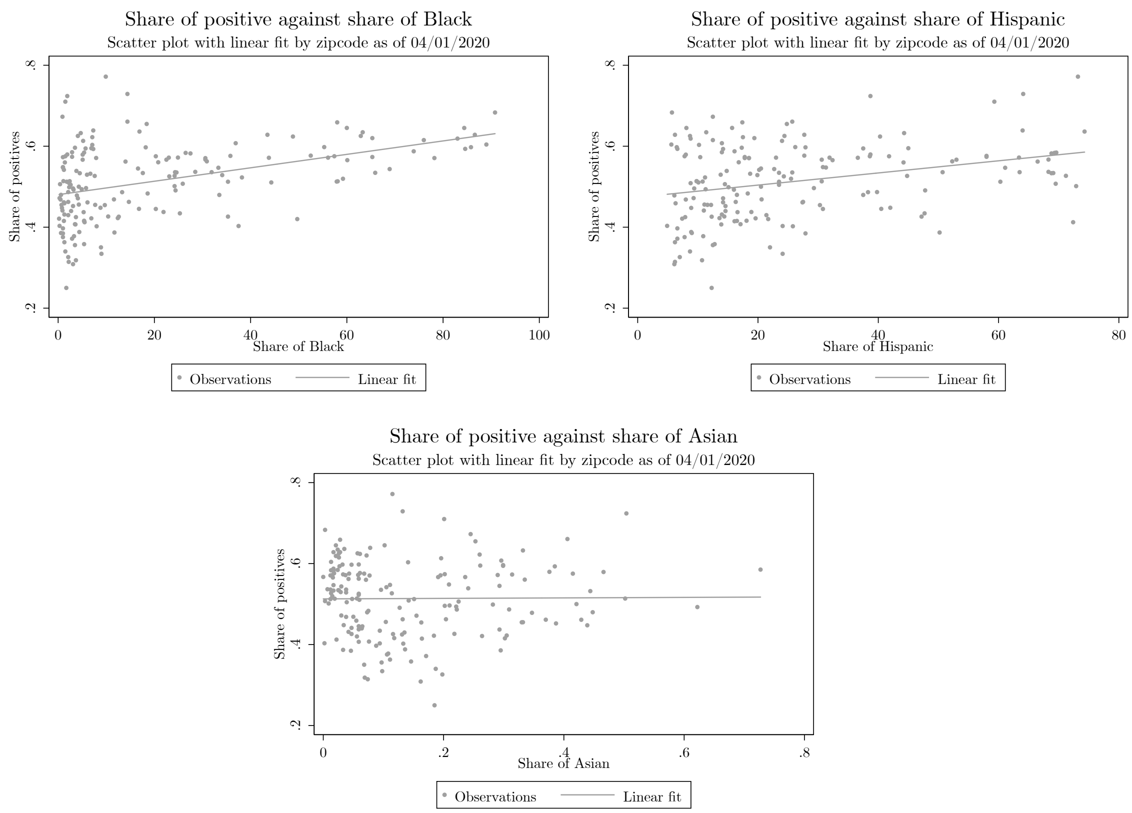

Figure 2: Share of positives against income, share of males, and household size by zipcode.

[10]These groups compose 29 % and 56 %, respectively, of all Bronx residents, 31 % and 19 % for Brooklyn, and 17 % and 28 % in the case of Queens.

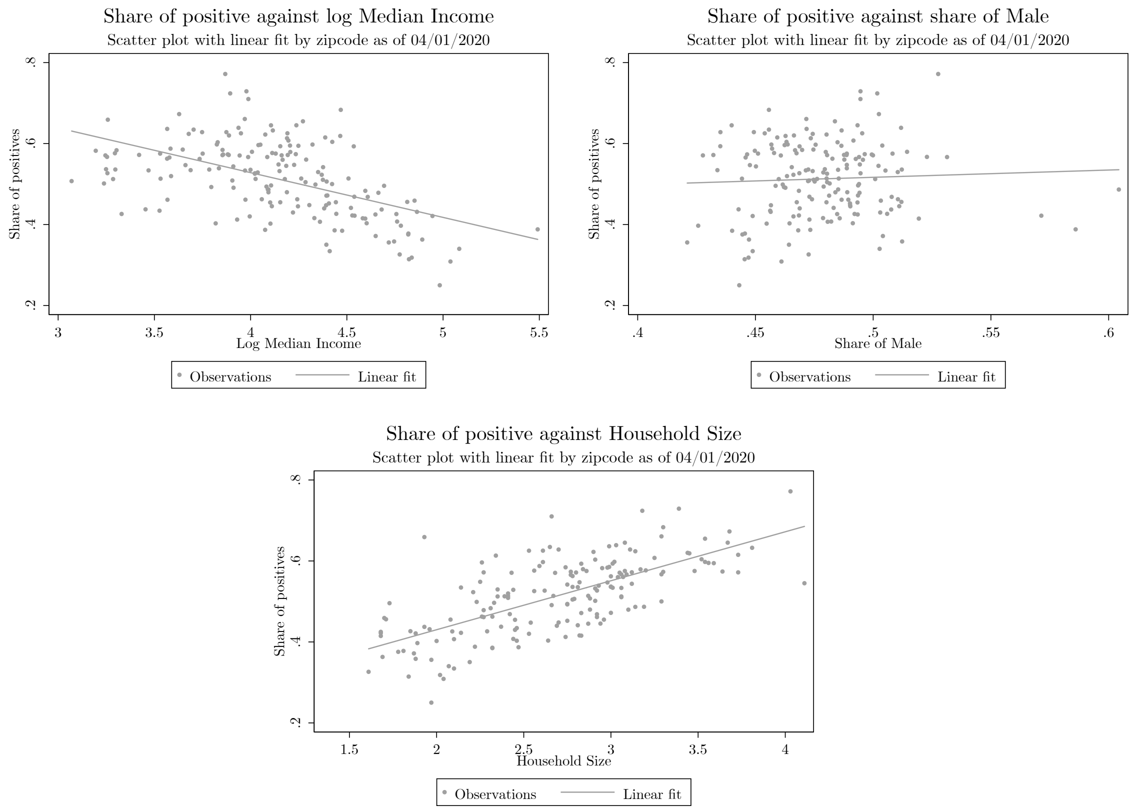

Figure 3: Share of positives against income, share of males, and household size by zipcode.

We also construct the shares of working age population employed at different occupations. These are categorized according to the groups listed in column 2 of Table 1 and the ACS provides the number of workers in each one of them, by zip code of residence. Taking the share of essential occupations into account, we divide the occupations into 13 categories for our empirical exercise, according to their essential definition, correlations between them and similarity in work environments and social exposure. Table 1 shows the occupation groups that will be used in the estimating regressions. Summary statistics for all variables included in our empirical analysis can be found in Table 2

Table 1: Occupation categories

Category	**ACS Occupations**
(1) Essential - Professional	Management, Business, Finance
(2) Non essential - Professional	Computer and Mathematical, Architecture and Engineering, Sales and Related, Community and Social Services, Education, Training, and Library, Arts, Design, Entertainment, Sports, and Media Administrative and Office Support
(3) Science fields	Life, Physical, and Social Science
(4) Law and related	Legal
(5) Health practitioners	Health practitioners
(6) Other health	Health technologists, technicians, and Healthcare Support
(7) Firefighting	Firefighting and prevention
(8) Law enforcement	Law enforcement
(9) Essential - Service	Food Preparation and Serving, Building and Grounds Cleaning and Maintenance
(10) Non essential - Service	Personal Care and Service
(11) Industrial, Natural resources and Construction	Construction and Extraction, Material Moving, Farming, Fishing, and Forestry, Production
(12) Essential - Technical	Installation, Maintenance, and Repair
(13) Transportation	Transportation

Finally, Table 2 presents the summary statistics of all the variables that are used in our analysis.

Table 2: Summary statistics

Variable	**Mean**	**Std. Dev.**	**p10**	**Median**	**p90**
Share of positives	0.563	0.085	0.438	0.583	0.645
Tests per Capita	0.018	0.006	0.012	0.017	0.026
Median Income (in 000's)	68.604	31.878	34.122	62.202	115.084
Share $\geq 20, \leq 40$	0.323	0.084	0.246	0.308	0.433
Share $\geq 40, \leq 60$	0.258	0.033	0.220	0.261	0.296
Share ≥ 60	0.200	0.079	0.132	0.190	0.276
Share Male	0.477	0.029	0.446	0.479	0.508
Household Size	2.683	0.537	1.930	2.750	3.300
% Black	0.200	0.240	0.010	0.076	0.600
% Hispanic	0.263	0.195	0.078	0.189	0.634
% Asian	0.144	0.139	0.017	0.094	0.335
Density (in 000's)	43.380	31.045	10.784	36.639	90.075
% Public Transport	0.532	0.150	0.312	0.543	0.712
Commuting Time (in mins)	40.647	7.054	27.200	42.100	48.100
% Uninsured	0.089	0.043	0.042	0.084	0.143
% Essential - Professional	0.126	0.089	0.046	0.092	0.285
% Non ess. - Professional	0.279	0.075	0.195	0.271	0.359
% Science fields	0.006	0.007	0.001	0.004	0.015
% Law and related	0.018	0.026	0.003	0.008	0.049
% Health practitioners	0.029	0.018	0.009	0.026	0.050
% Other health	0.038	0.024	0.010	0.035	0.073
% Firefighting	0.012	0.009	0.003	0.012	0.023
% Law enforcement	0.007	0.007	0.001	0.006	0.014
% Essential - Service	0.065	0.033	0.035	0.060	0.107
% Non ess. - Service	0.032	0.013	0.016	0.032	0.047
% Ind. and Construction	0.054	0.027	0.014	0.056	0.090
% Essential - Technical	0.014	0.009	0.004	0.013	0.022
% Transportation	0.029	0.016	0.004	0.032	0.048

3 Results

3.1 General Results

We present the main empirical results in this section, for our four different specifications. Our unit of analysis is the zip code and all models include the share of positive tests as the dependent variable. Additionally, we include tests per capita to control for selection on testing. The first model includes some widely discussed potential factors of COVID spread in NYC, density and commuting patterns — specifically, log of population density, percentage of workers using public transport, and average commuting time. We also include the percentage of uninsured as a health control. Our second model expands by including our proposed mechanism, the percentage of working age population employed in each of the 13 occupation categories defined in Table 1. The third specification adds demographic controls related to income, age, gender, household size, and race. Finally, we include borough fixed effects in our last model. Exploiting the fact that we have daily data over multiple days, we estimate a separate regression for each one of them, allowing us to detect any time variation in the effects. The daily evolution is analyzed in the next section.

The first model shows the effect of the variables commonly used to explain COVID-19 incidence in NYC. While Harries (2020) finds that subway use was a major factor of the virus spread, we find that it does not have a significant effect. This result could be due to the there is not enough cross neighborhood variation to identify this effect as most New Yorkers use public transport as their daily commute. Nonetheless, commuting time is a significant factor. For example, for April 1 an increase in commuting time of four minutes, a 10% increase, correlates with an 2.3% increase in the share of positives. We also find a positive and significant effect of the share of uninsured population on the rare of positives. This result can be explained by patient selection, with uninsured patients only being willing to be tested under very acute symptoms in fear of medical charges. While its magnitude decreases as we include other covariates, the estimate of this variable remains positive and significant with a one percentage increase in the share of uninsured population being correlated with a 0.65% increase in the rate of positives for an average of 54% rate of positives as of April 20, 2020.

Specification (2) shows the effect of different occupations in share of positives. The variables are defined as shares of the working age population employed in these occupations, and so the coefficients are relative to the working age but not employed. The coefficients can be read as the effect in percentage points on the positive share of an increase in one percentage point in the population employed in the particular category. We find that some occupations can significantly explain part of the variation in COVID-19 incidence. On the one hand, an increase in the share of workers employed in non essential - professional, other health (not health practitioners), and transportation occupations are all associated with a higher percentage of positive tests. On the other hand, higher shares of workers in science fields, legal occupations, and law enforcement have a negative effect of share of positives. These results will be further discussed in the time trends section.

Perhaps surprisingly, under this specification commuting time no longer has a significant effect. This result suggests that commuting patterns are closely related to occupations and most of the explanatory variation for commuting patterns comes through this channel. This result also implies the existence of within city location and mobility patterns that are occupation specific.

We include demographic variables in the third model. Despite the strong correlation between share of positives and demographic characteristics, the results for specification (3) show that some of it can be explained through the occupation mechanism. Notably, the income effect disappears when controlling for occupations, suggesting that the previous correlation is due to income differences across jobs. Still, some demographic effects remain significant. For April 20, a one percentage point increase for Blacks and Hispanics leads to 0.15% and 0.23% in the rate of positive respectively, an effect that is economically small. A plausible explanation for these patterns could be driven by a racial bias on the incidence of testing as pointed out by Borjas (2020). Another explanation is differences in adherence to the shelter-in-place policy, as explored by Coven and Gupta (2020). We also find that household size has an effect on tests. Adding one extra person to the average household, a 37% increase, leads to a 7% increase in the percentage of positive tests. While neighborhood density does not explain variation in share of positives, density in households appears to do so, with a an effect that increases over time.

The tests per capita coefficient is positive and highly significant across most specifications, suggesting there is selection in testing. With the scarcity of tests in the past weeks, those tested in NYC were selected from a population with more acute symptoms and therefore a higher probability of being positive. If only those who at high risk of having the disease are tested, more tests should unambiguously lead to higher rates of positive results.[11] While its magnitude decreases over time, when including borough fixed effects it remains consistently significant. This type of pattern is consistent with more aggressive selection for the boroughs of Bronx and Staten Island, or with more widespread testing for the boroughs of Brooklyn and Queens, all relative to Manhattan.

[11] Another type of selection could be driven by higher income households being able to pay for more testing. In this case, we should expect a negative sign. However, given the way testing has been managed in NYC (only testing after hospitalization at first with the introduction of free testing later on), we believe that this type of selection is unlikely.

Table 3: Dependent variable - Share of Positives as of April 1, 2020

	(1) Nbhd Controls		(2) + Occupations		(3) + Demographics		(4) + Borough FE	
Tests per capita	9.017***	(2.879)	11.186***	(2.447)	10.773***	(2.249)	12.050***	(2.386)
Log Density	0.015	(0.014)	0.022*	(0.012)	0.015	(0.012)	0.032***	(0.011)
% Public Transport	-0.015	(0.072)	0.013	(0.068)	0.053	(0.070)	-0.059	(0.062)
Log Commuting Time	0.237***	(0.046)	-0.016	(0.083)	0.009	(0.075)	-0.054	(0.062)
% Uninsured	1.002***	(0.141)	0.662***	(0.246)	0.336	(0.215)	0.150	(0.180)
% Essential - Professional			0.156	(0.271)	0.695***	(0.238)	0.766***	(0.236)
% Non ess. - Professional			0.669***	(0.189)	0.615***	(0.181)	0.544**	(0.216)
% Science fields			-4.703***	(1.294)	-3.745***	(1.064)	-2.965***	(1.118)
% Law and related			-0.410	(0.801)	-0.875	(0.754)	-1.427**	(0.697)
% Health practitioners			-0.432	(0.421)	-0.167	(0.431)	-0.167	(0.386)
% Other health			0.947***	(0.321)	0.027	(0.412)	0.346	(0.402)
% Firefighting			2.743**	(1.072)	1.624	(1.109)	1.629*	(0.965)
% Law enforcement			-0.301	(1.215)	0.815	(1.089)	-0.223	(1.016)
% Essential - Service			-0.100	(0.354)	0.258	(0.347)	0.245	(0.300)
% Non ess. - Service			0.769	(0.561)	1.166**	(0.509)	1.154**	(0.483)
% Ind. and Construction			1.091**	(0.437)	1.208***	(0.402)	0.839**	(0.401)
% Essential - Technical			-2.025*	(1.133)	-0.457	(0.979)	-0.319	(0.881)
% Transportation			1.752***	(0.588)	1.718***	(0.527)	1.102**	(0.469)
Log Income					-0.008	(0.034)	-0.010	(0.033)
Share $\geq$ 20, $\leq$ 40					-0.346*	(0.176)	-0.357**	(0.173)
Share $\geq$ 40, $\leq$ 60					-0.855***	(0.222)	-0.611**	(0.237)
Share $\geq$ 60					-0.380**	(0.175)	-0.347*	(0.197)
Share Male					-0.050	(0.267)	-0.146	(0.264)
Log Household Size					0.076	(0.073)	0.037	(0.061)
% Black					0.149***	(0.039)	0.175***	(0.040)
% Hispanic					0.003	(0.050)	0.194***	(0.050)
% Asian					0.136**	(0.053)	0.141***	(0.050)
Bronx							-0.014	(0.023)
Brooklyn							0.086***	(0.022)
Queens							0.084***	(0.024)
Staten Island							0.083***	(0.027)
Constant	-0.671**	(0.264)	-0.149	(0.342)	0.110	(0.372)	0.196	(0.334)
Observations	174		174		174		174	
R^2	0.514		0.694		0.785		0.839	

Weighted OLS by population size. Robust standard errors in parentheses

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

Table 4: Dependent variable - Share of Positives as of April 10, 2020

	(1) Nbhd Controls		(2) + Occupations		(3) + Demographics		(4) + Borough FE	
Tests per Capita	1.913**	(0.832)	1.713*	(0.921)	1.795**	(0.763)	3.904***	(0.675)
Log Density	0.022*	(0.013)	0.018*	(0.010)	0.013	(0.008)	0.022***	(0.007)
% Public Transport	-0.001	(0.060)	0.012	(0.056)	0.095*	(0.056)	-0.004	(0.041)
Log Commuting Time	0.262***	(0.040)	0.019	(0.069)	0.022	(0.060)	-0.023	(0.045)
% Uninsured	1.038***	(0.104)	0.521***	(0.187)	0.316**	(0.136)	0.290***	(0.103)
% Essential - Professional			-0.003	(0.203)	0.579***	(0.177)	0.484***	(0.179)
% Non ess. - Professional			0.419**	(0.173)	0.354**	(0.166)	0.257*	(0.138)
% Science fields			-3.021***	(1.082)	-3.094***	(0.905)	-2.334***	(0.812)
% Law and related			-0.604	(0.525)	-1.050**	(0.480)	-1.293***	(0.422)
% Health practitioners			-0.248	(0.372)	-0.061	(0.379)	-0.124	(0.281)
% Other health			0.753***	(0.258)	-0.275	(0.302)	0.238	(0.231)
% Firefighting			1.282	(0.880)	-0.042	(0.869)	0.456	(0.570)
% Law enforcement			-1.859	(1.149)	-1.217	(0.892)	-1.323*	(0.751)
% Essential - Service			0.159	(0.262)	0.198	(0.280)	0.127	(0.199)
% Non ess. - Service			0.359	(0.471)	0.844**	(0.417)	0.781**	(0.350)
% Ind. and Construction			0.472	(0.332)	0.497*	(0.279)	0.101	(0.225)
% Essential - Technical			-0.729	(0.854)	-0.150	(0.718)	-0.474	(0.531)
% Transportation			1.824***	(0.419)	1.639***	(0.377)	0.831***	(0.299)
Log Income					-0.024	(0.023)	-0.027	(0.021)
Share $\geq$ 20, $\leq$ 40					-0.243**	(0.122)	-0.246**	(0.097)
Share $\geq$ 40, $\leq$ 60					-0.510***	(0.181)	-0.228	(0.149)
Share $\geq$ 60					0.127	(0.114)	-0.017	(0.115)
Share Male					0.453**	(0.180)	0.249	(0.171)
Log Household Size					0.167***	(0.054)	0.111***	(0.042)
% Black					0.165***	(0.030)	0.114***	(0.026)
% Hispanic					0.018	(0.041)	0.130***	(0.034)
% Asian					0.047	(0.043)	0.018	(0.030)
Bronx							-0.040***	(0.014)
Brooklyn							0.053***	(0.015)
Queens							0.058***	(0.016)
Staten Island							-0.022	(0.023)
Constant	-0.765***	(0.228)	0.010	(0.269)	-0.137	(0.258)	0.107	(0.213)
Observations	174		174		174		174	
R^2	0.674		0.801		0.871		0.921	

Weighted OLS by population size. Robust standard errors in parentheses

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

Table 5: Dependent variable - Share of Positives as of April 20, 2020

	(1) Nbhd Controls		(2) + Occupations		(3) + Demographics		(4) + Borough FE	
Tests per Capita	0.667	(0.485)	0.262	(0.560)	0.381	(0.497)	2.553***	(0.476)
Log Density	0.024**	(0.011)	0.015*	(0.009)	0.011	(0.008)	0.016***	(0.006)
% Public Transport	0.010	(0.055)	-0.001	(0.050)	0.080	(0.053)	-0.017	(0.040)
Log Commuting Time	0.232***	(0.034)	0.001	(0.060)	0.004	(0.055)	-0.008	(0.042)
% Uninsured	0.924***	(0.099)	0.417**	(0.171)	0.296**	(0.129)	0.351***	(0.098)
% Essential - Professional			-0.210	(0.168)	0.294*	(0.165)	0.235	(0.160)
% Non ess. - Professional			0.329**	(0.147)	0.274*	(0.152)	0.224*	(0.122)
% Science fields			-1.931*	(1.016)	-2.318***	(0.861)	-1.609**	(0.784)
% Law and related			-0.492	(0.456)	-0.851*	(0.460)	-0.898**	(0.397)
% Health practitioners			-0.155	(0.357)	0.010	(0.387)	-0.206	(0.278)
% Other health			0.815***	(0.232)	-0.053	(0.272)	0.365	(0.221)
% Firefighting			0.379	(0.829)	-0.765	(0.876)	-0.156	(0.556)
% Law enforcement			-1.970*	(1.049)	-1.472*	(0.820)	-1.344**	(0.655)
% Essential - Service			0.312	(0.229)	0.205	(0.242)	0.082	(0.171)
% Non ess. - Service			-0.046	(0.437)	0.455	(0.378)	0.578*	(0.296)
% Ind. and Construction			0.271	(0.317)	0.246	(0.271)	-0.079	(0.209)
% Essential - Technical			-0.785	(0.724)	-0.603	(0.617)	-0.908*	(0.487)
% Transportation			1.253***	(0.364)	1.083***	(0.327)	0.541*	(0.293)
Log Income					-0.021	(0.022)	-0.022	(0.019)
Share $\geq$ 20, $\leq$ 40					-0.169	(0.115)	-0.208**	(0.090)
Share $\geq$ 40, $\leq$ 60					-0.389**	(0.161)	-0.198	(0.126)
Share $\geq$ 60					0.248**	(0.108)	0.002	(0.104)
Share Male					0.540***	(0.166)	0.318**	(0.150)
Log Household Size					0.167***	(0.047)	0.099***	(0.036)
% Black					0.140***	(0.030)	0.081***	(0.026)
% Hispanic					0.027	(0.036)	0.125***	(0.033)
% Asian					0.015	(0.043)	0.012	(0.031)
Bronx							-0.062***	(0.014)
Brooklyn							0.034**	(0.014)
Queens							0.023	(0.015)
Staten Island							-0.064***	(0.022)
Constant	-0.682***	(0.195)	0.201	(0.232)	-0.076	(0.236)	0.111	(0.196)
Observations	174		174		174		174	
R^2	0.673		0.800		0.866		0.920	

Weighted OLS by population size. Robust standard errors in parentheses

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

3.2 Daily comparison and time trends

We find some time-variant results that could provide insight on both the evolution of the pandemic effects as well as the health policies that have been set in place. Figures 4 to 6 show the time evolution of the coefficients for specification (4). The result for the tests per capita variable is particularly salient; we observe a very positive effect on the share of positive tests that becomes progressively smaller over time. This result could be reconciled with the fact that in the earlier days of the crisis, testing was severely limited. Zip codes with more tests implied a higher share of people at high risk of having the disease. So, a key takeaway from the results of our daily comparison is the importance of widespread testing, as it allows us to identify the mechanisms that explain demographic and occupational differences in COVID-19 exposure.

There are notable time trends in the effects associated with occupations. Higher shares of professional and non essential - service categories were associated with higher percentage points in the rate of positive tests. An additional percentage point in either occupation group implied an almost 0.6 percentage point increase in the positive rate. However, they eventually turn not significant or trend towards zero, averaging closer to a 0.3 percentage point increase effect. A plausible explanation for this is that these professions are either non essential, or have the highest shares of remote workers. While highly exposed to the virus in the beginning, once the workers stay in place their effect of positive tests subsides. The opposite happens with science field and law occupations — they are negatively correlated with COVID-19 incidence in the beginning but the effect trends towards zero.

There are interesting patterns for the essential occupations as well. An additional percentage point in the share of transportation workers is associated with between 0.5 and a 1 percentage point increase in the positive rate. The effect seems to decay over time, but at a slower rate than other occupations. This result could be due to its essential designation, but also due to its relatively high exposure nature. The share of industrial, natural resources, and construction occupations starts off with a positive effect on COVID-19 incidence. However, construction was determined to be not essential a week after the general stay-at-home order, and this order could explain its eventual attenuation. Law enforcement occupation shares have a consistently negative effect on positive shares, while firefighter shares have declining trajectory towards zero. A plausible explanation for this difference could be the partnership between NYPD and health care groups to provide free testing to its members.[12]

The share of uninsured population increasingly predicts the variation in positive tests. We find that an additional percentage point of uninsured predicts an almost 0.3 point increase in the share of positives. While many health care providers are waiving COVID-19 related out of pocket costs, these remain very high for those uninsured and so a higher incidence of COVID-19 in this group could imply a severe financial burden. While still significant, the effect of neighborhood density declines over time,

[12] www.nypost.com/2020/04/03/nypd-partners-with-health-care-groups-to-test-cops-for-covid-19/

and the opposite occurs for household size. The stay-at-home order could mitigate part of the risk of high neighborhood density, while increasing the probability of within-household infections.

Finally, another outstanding time pattern is that the coefficients on racial composition decrease in magnitude as the selection of testing decreases. This result may suggest that there could be a stronger racial selection component among those in worse conditions at earlier dates. For example, an explanation for this pattern could be that Black citizens were less likely to be tested or had to be in worse conditions to access testing compared to White citizens.[13]

4 Conclusions and policy implications

In this paper, we present evidence showing that occupations are an important channel in explaining the difference in rates of COVID-19 across neighborhoods. Using data from NYC at the zip code level, we study the relationship between the share of positive tests and the share of workers in different occupations. The DOH provides daily updates of COVID-19 test data, allowing us to study the aforementioned relationships over multiple days and detect time variation in their magnitudes.

We begin by showing descriptive evidence of heterogeneous incidence of positive cases across neighborhoods, income, race, gender, and household size. A zip code's median income is negatively correlated with its share of positives. Conversely, we find a positive correlation between shares of Black and Hispanic residents as well as average household size with the share of positive tests. Highlighting these differences is important as it confirms that the disease has had more harmful effects on vulnerable communities and finding an occupation mechanism that explains it could guide policy intended to alleviate its impact.

We estimate several models to explore the effects of occupations. Our first specification only includes neighborhood characteristics including the use of public transportation and the average length of daily commutes. While it has been argued that commuting patterns are major factor in the spread of the disease in NYC, we show that after including occupation controls they fail to significantly explain variation in share of positives.

We then augment the model by including the occupation shares. We find the strongest positive effect on share of positive tests in the share of workers on Transportation, Industrial, Natural resources, Construction, and Non essential - Professional, with significant time variation in it. For example, in the case of Transportation a point increase in the percentage of workers in these occupations has an effect on positive rates of between 2 and 1 percentage points throughout the period. While the other two have a significant effect in positive shares in the first days, their effect

[13]Some evidence that this is plausible mechanism can be found in www.modernhealthcare.com/safety-quality/long-standing-racial-and-income-disparities-seen-creeping-covid-19-care

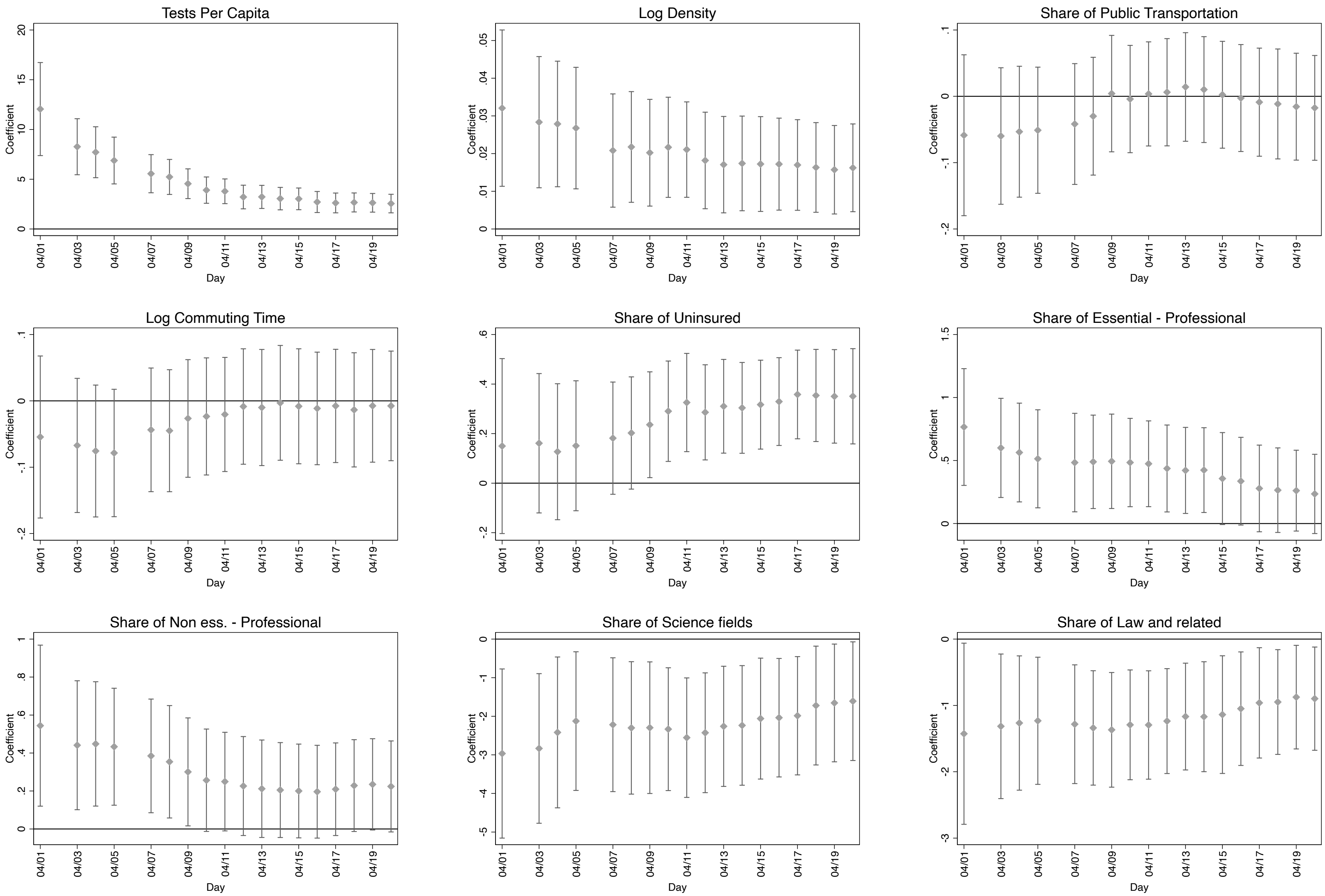

Figure 4: Regression coefficients of specification (4) over time

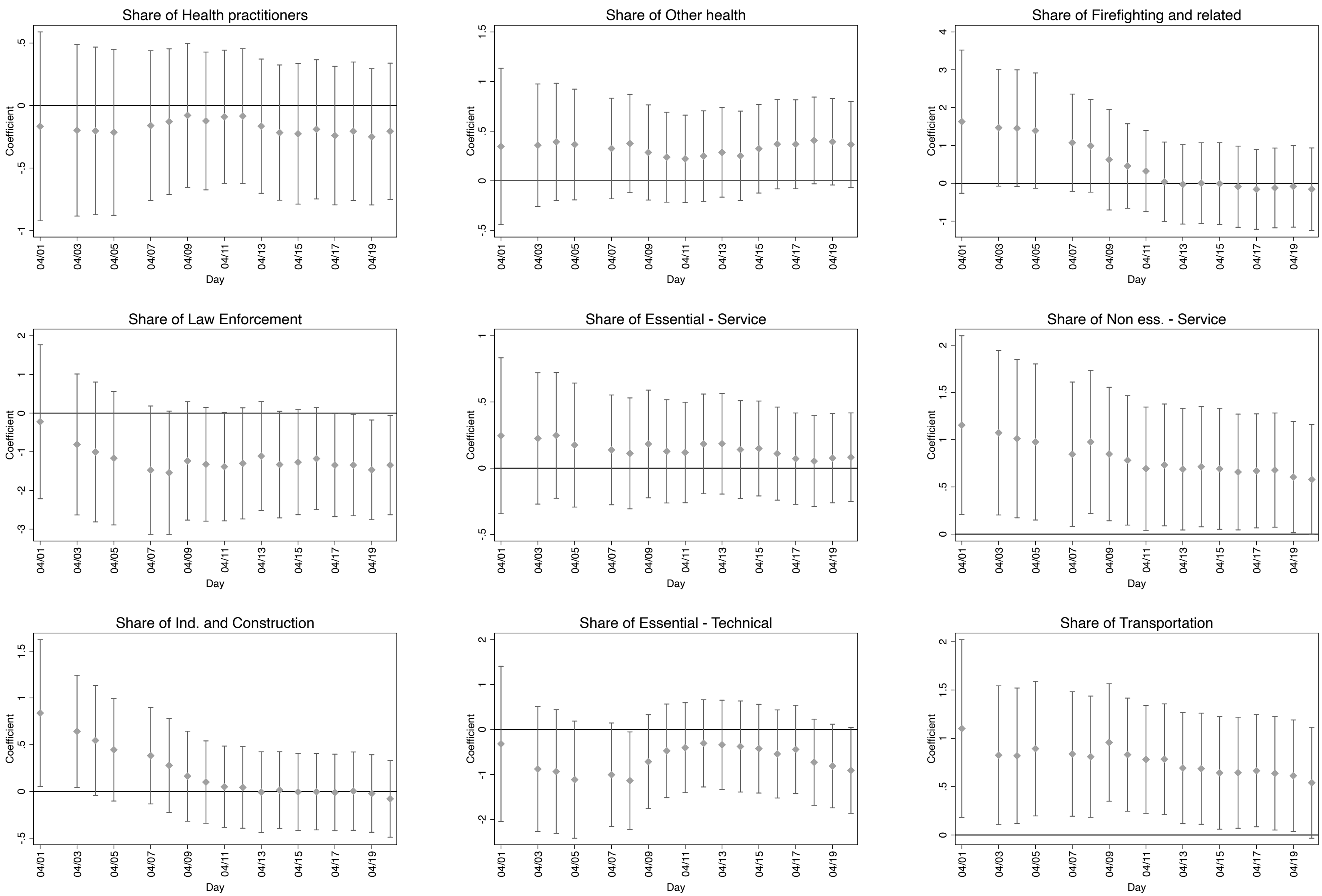

Figure 5: Regression coefficients of specification (4) over time

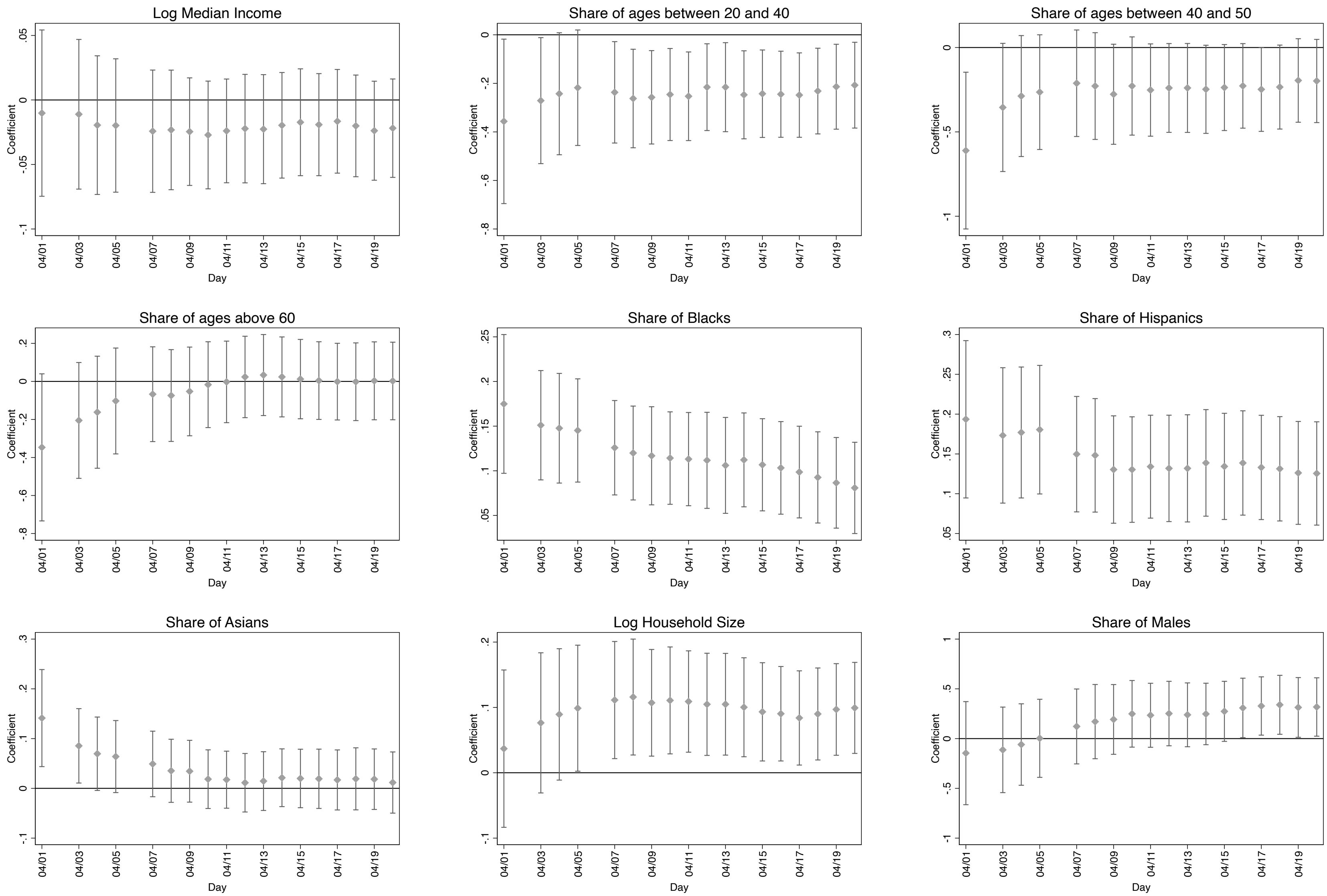

Figure 6: Regression coefficients of specification (4) over time

becomes insignificant by the end of the sample period. This could be an effect of the stay-at-home order. Conversely, higher shares of workers in science fields and law enforcement have a negative effect on positive tests, with science fields decreasing its magnitude over the days.

When adding demographic controls we observe that racial effects do persist, suggesting that the occupation mechanism does not fully explain the entire racial disparity. However, their magnitude is small, and arguably not economically relevant when compared to the statistically significance of the occupation effects. Income and most age groups do not contribute to explain the variation in positive tests, suggesting that disparities along those demographics observed in the data can be explained through the occupation mechanism.

In all of our regression models we include a test per capita, finding that it is a strong predictor of share of positive tests. However, its relative importance declines each passing day. This trend suggests that there was significant selection early on, but as tests become more available, this issue loses relevancy and more of the variation in COVID-19 incidence can be explained through the occupation channel.

Our results suggest clear implications for policy all over the world, but in particular for NYC. First, as mentioned earlier, our results highlight the need of mass testing to be able to cleanly identify what are the relevant channels that generate a greater risk of exposure. Second, once these channels are identified policy makers can target specific groups in the provision of protective gear, tests, and vaccinations. We think of the purpose of this policy as twofold. On the one hand, it provides a form of insurance against contracting the disease to those who are more vulnerable. On the other hand, targeting specific workers or demographic groups may have spillovers on the rest of the population. For example, a policy that starts vaccinating and/or testing those workers with higher rates of human interactions does not only have effects on those directly targeted by the policy but also on those who are likely to be in contact with them. We believe, given the high contagion rates of COVID-19, that these type of spillovers from this policy may be substantial. Moreover, given that access to insurance plays a significant role that becomes more important over time, another policy recommendation is to incentive testing for those without medical insurance such as a complete coverage of out-of-pocket costs in relation to COVID-19 including testing. In this way, the authorities would be able to have a clearer image of potential focal points of contagion. Finally, our results also suggest that that stay-at-home order could have mitigated contagion rates at work or in public spaces, while increasing the probability of intra-household infections. This contagion within a single home could be alleviated by a policy offering shelter to those families with a large number of members or with a limited sharing space.

References

Barbieri, T., Basso, G., and Scicchitano, S. (2020). Italian workers at risk during the covid-19 epidemic. *Unpublished Manuscript.*

Borjas, G. J. (2020). Demographic determinants of testing incidence and covid-19 infections in new york city neighbourhoods. *Covid Economics, Vetted and Real-Time Papers.*

Coven, J. and Gupta, A. (2020). Disparities in mobility responses to covid-19. *Working Paper.*

Furth, S. (2020). Automobiles seeded the massive coronavirus epidemic in new york city. Available at `https://marketurbanism.com/2020/04/19/automobiles-seeded-the-massive-coronavirus-epidemic-in-new-york-city/`.

Harries, J. E. (2020). The subways seeded the massive coronavirus epidemic in new york city. *NBER Working Papers.*

Schmitt-Grohé, S., Teoh, K., and Uribe, M. (2020). Covid-19: Testing inequality in new york city. *NBER Working Papers.*