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

Milena Almagro† Angelo Orane-Hutchinson‡

April 29, 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 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 no longer 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, although their magnitudes are economically small. Additionally, we perform the same analysis over a time window to evaluate how different channels interact with the progression of the pandemic, as well as with the health policies that have been set in place. While the coefficient magnitudes of many occupations and demographics decrease over time, we find evidence consistent with higher intra-household contagion as days go by. Moreover, our findings also suggest a selection on testing, whereby those residents in worse conditions are more likely to get tested, with such selection decreasing over time as tests become more widely available.

*We thank Michael Dickstein and Daniel Waldinger for their useful comments. 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

The impact of COVID-19 has affected different locations to very different extents, with some areas being hit harder than others all over the world. Much of this variation is explained by characteristics such as the number of international travellers, weather conditions, local policies to control the pandemic, and when those policies were implemented. Surprisingly, large differences exist even across smaller geographical units such as neighborhoods *within* a city. For example, Figure 1 shows the differences in the rates of positive tests by zip code of residence in New York City (NYC).

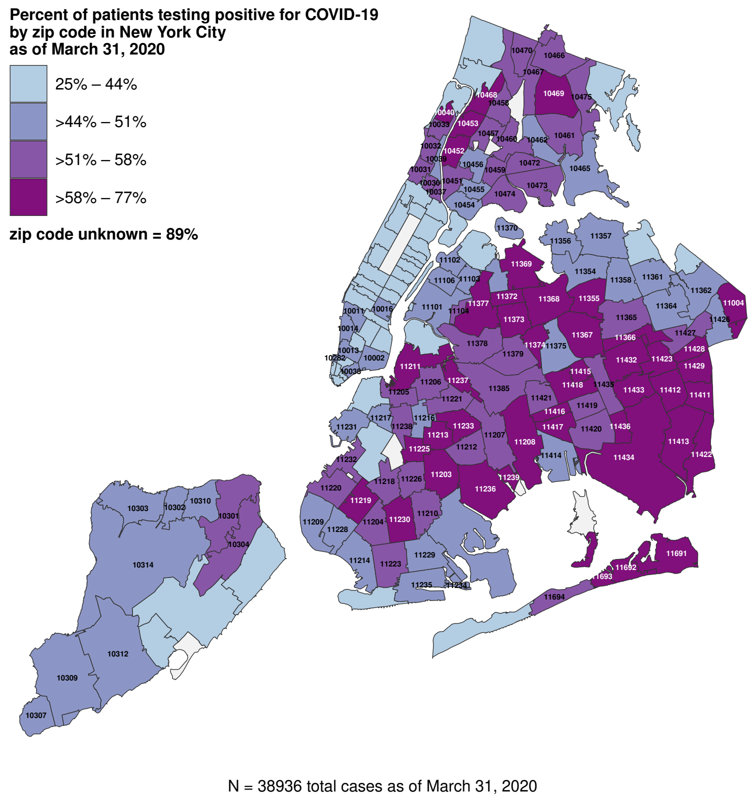

Figure 1: Map of the rate of positives by zip code as of March 31, 2020.

From simple inspection, zip codes with the highest rates are found in the boroughs of Bronx, Brooklyn, and Queens. These boroughs are also home to the majority of Blacks and Hispanics living in NYC.[1]

These spatial correlations between the incidence of the pandemic and demographics have garnered the attention of many economists and policy makers. For example,

[1] 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.

Borjas (2020) and Schmitt-Grohé et al. (2020) show that the much of the spatial disparities of testing and positive rates across NYC neighborhoods is explained by demographics. Given that COVID-19 does not intrinsically discriminate across demographic groups, the reason for such disparities still remains an open question. Hence, our goal is to assess the importance of a set of observable channels, such as population density, commuting patterns, and occupations, that explain the existing spatial differences in NYC.

To understand the relevance of different mechanisms, we use data on the number of tests and positives across NYC zip codes provided by DOH.[2] Because these data have been released (almost) on a daily basis, we are able to keep track of the number of tests and the fraction of those that are positive since April 1. We combine the data on testing with neighborhood and demographic indicators, which are provided by the American Community Survey (ACS). Namely, we use zip code level data on population density, commuting patterns, income, as well as race and age composition. We also include employment data; the ACS provides the number of workers employed at different occupations, all at the zip code level. We compute the share of workers across different occupations relative to the working-age population to understand how differences in labor composition can affect the incidence of COVID-19.

Because we focus on highlighting observable channels that are likely to explain the spatial differences to COVID-19 exposure, we estimate several specifications highlighting the importance of new variables at each step. Throughout the analysis, our dependent variable is the fraction of tests showing a positive result across NYC zip codes.[3] 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 share of the population being tested, which we call "tests-per-capita." The limited availability of tests in NYC has forced health authorities to constrain testing to people showing sufficiently acute symptoms or determined to be at high risk of infection. Hence, we expect the number of tests administered to be very close to the population in that segment.[4] Therefore, we use the number of tests per capita as a proxy for the overall level of the spread of the pandemic *within* a neighborhood. We find that when the number of tests per capita increases, the share of positive tests also increases. This result stems from both variables co-moving with the true

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

[3]We could also focus on the number of positive tests per capita. We refrain from doing so for two reasons. First, random testing has not been possible in NYC, as only those with certain conditions are tested because of limited testing capacity. Second, Borjas (2020) points out that the incidence of different variables on positive results per capita is composed of two things: A differential incidence on those who are tested, but also a differential incidence on those with a positive result conditional on being tested. Therefore, we believe that the fraction of positive tests is the variable that correlates the most with the actual spread of the disease within a neighborhood throughout our sample.

[4]As a matter of fact, at earlier dates, tests were performed only on those who required hospitalization.

number of infected people within a neighborhood. However, we also find that, as testing becomes more widely available and more tests are performed on the asymptomatic population, the magnitude of tests-per-capita decreases over our analyzed time period.

We then analyze the role of occupations, motivated by the fact that they vary in their degree of human interaction. Those with high levels of human contact are more likely to be exposed to the virus.[5] We do so by including the share of workers for 13 categories in each zip code constructed from the ACS according to their degree of human interaction. The results show 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 a one-percentage-point increase in the number of workers employed in transportation, an occupation that has been declared essential and has a high degree of exposure to human interaction, increases the share of positive tests by 2% for April 1, one month into the pandemic. Moreover, we show that after controlling for occupations, length of commute and the use of public transport are not significant.[6]

Additionally, these results are robust to the inclusion of demographics, as well as borough fixed effects.[7] Including demographics leads to several striking patterns. Whereas simple correlations show that wealthier neighborhoods have a lower rate of positives, we show that income is not significant when occupations are included. However, we still see significant and positive effects on positive rates for minorities. These results could be because minorities are less likely to get tested, or have to be in worse conditions than whites in order to get tested.[8] However, it can be questioned whether these racial disparities are economically relevant. Moreover, their magnitudes decrease over time as more testing becomes available – with Asians showing no statistical significance at the end of our sample. For example, on April 1, one month after the pandemic started in NYC, we find that a one-percentage point increase in the share of Blacks correlates with an increase of 0.34% in the share of positive tests, for an average number of 51% of positive cases. By April 20th, these numbers are 0.15% and 54%, respectively. For Hispanics, the disparity is larger, where a one-percentage-point increase in their population corresponds to an increase of 0.38% and 0.23% in the rate of positives, for the same two dates.

Our daily analysis also reveals that, as the stay-at-home orders starts to be effective, the magnitude of many occupations decrease as days go by. For example, a one-percentage-point increase in the number of workers employed in transportation

[5]Michaels et al. (2019) show that interactive occupations have become more important in larger metros over time. A recent paper by Barbieri et al. (2020) shows evidence of this mechanism for workers in Italy.

[6]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."

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

[8]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`

decreases its size to 1% as of April 20, almost two months into the pandemic and one month after the stay-at-home order went into effect. On the other hand, we still find a rather stable coefficient of household size over time, which is consistent with the stay-at-home order being more helpful at mitigating contagion at work or in public spaces than within the household.

We conclude that much of the disparity in the rates of positives can be explained by different demographic groups being more or less representative across different occupations. In particular, a key channel appears to be the differences in exposure to human contact across jobs. However, our results also suggest that the relevance of these variables decreases over time, and this change is accompanied by an increase in intra-household contagion as days go by. These trends are consistent with the progression of the pandemic and its interaction with the policies set in place. Two immediate policy implications arise from our analysis. First, it would be desirable to target these more sensitive groups of occupations with the distribution of protective gear, testing, and vaccination. This policy should not only be considered for their own risk of exposure, but also for the risk to others due to potential spillovers on the rest of the population. Second, local governments could give access to temporary shelter to those households that are forced to live in a reduced shared space.

2 Data description and patterns

Our source of incidence rates of COVID-19 and the number of tests performed is the NYC DOH data release. The DOH releases (almost) daily data on the cumulative count of COVID-19 cases and the total number of residents that have been tested, divided by the zip code of residence. We have collected data from April 1 to April 24, with only April 2 and April 6 missing from our sample.[9]

We obtain demographic and occupation data at the zip code level from the ACS. The demographic characteristics we include are zip code median income, average age, racial breakdown, and health insurance status. We also include commuting-related variables: average commute time to work as well as means of transportation. We plot a simple correlation between the share of positive tests and demographics. We see that shares of Blacks and Hispanics are positively correlated with rate of positive tests, a flat relationship for the share of Asians, and a negative relationship for income, as shown in Figure 2.

[9] Unfortunately these days have never been made publicly available.

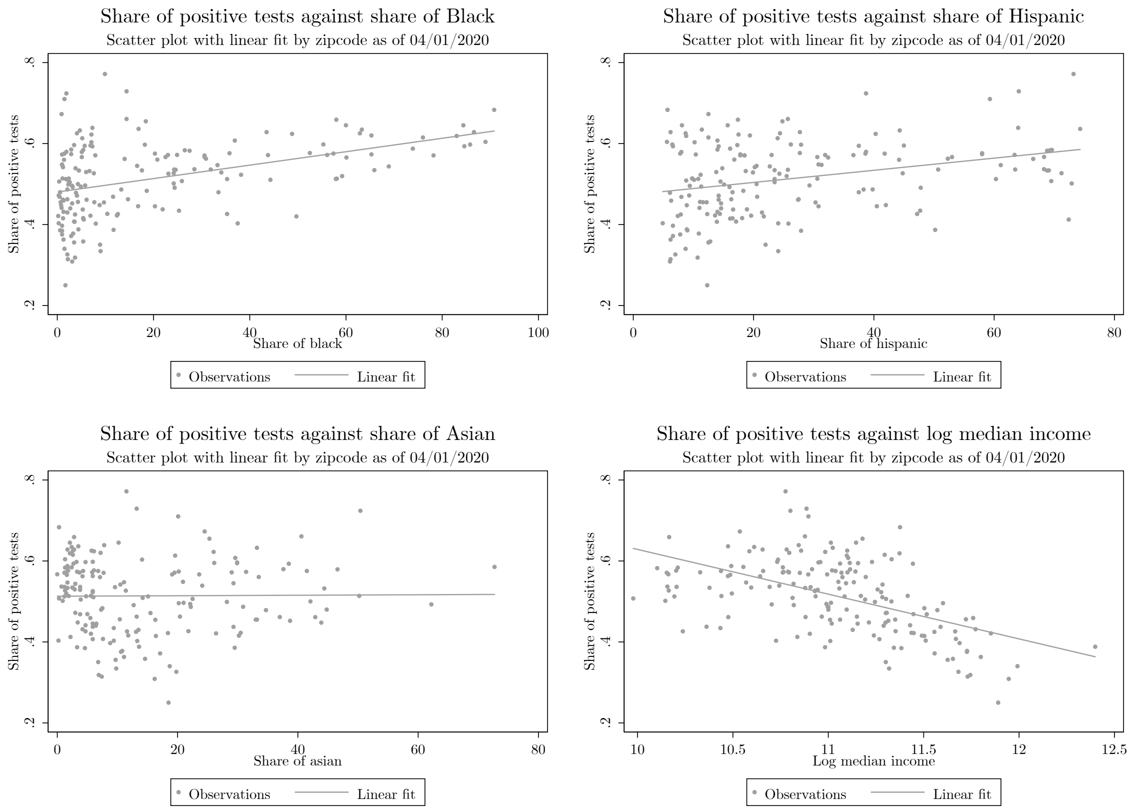

Figure 2: Share of positive tests against demographics by zip code

We also construct the shares of the working-age population employed at different occupation categories. The ACS provides the number of workers employed in each of the occupations listed in column 2 of Table 1 by zip code of residence. We then categorize them according to the groups listed in column 1 of Table 1. We do so by taking into account their essential definition, spatial correlations between them, and similarity in work environments and social exposure.[10] Table 1 shows the occupation groups that we use in our regressions. Summary statistics for all variables included

[10]Leibovici et al. (2020) rank occupations according to an index of occupational contact-intensity, defined from a survey by O*NET. They use ACS individual-level data at the four digit Standard Occupation Classification (SOC) level and match it to 107 ACS-defined occupations. Unfortunately, we only observe occupations at the SOC first level of aggregation for zip code data and cannot match their classification to our spatial distribution. Nonetheless, our categorization closely follows the intensity index grouping for the more specific group of occupations when aggregated to the first SOC level. More importantly, when defining our 13 categories, we avoid mixing occupations with large differences in their contact-intensity values. For robustness we have also performed our analysis with two alternative classifications for occupations. First, we divided occupations between essential and non-essential as declared by the US government. Second, we used the four categories defined in Kaplan et al. (2020). In both cases, the high level of aggregation lead to non-significant estimates or results that were hard to reconcile with observational evidence.

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 positive tests	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 ≥ 20, ≤ 40	0.323	0.084	0.246	0.308	0.433
Share ≥ 40, ≤ 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
% Essential - Service	0.065	0.033	0.035	0.060	0.107
% Essential - Technical	0.014	0.009	0.004	0.013	0.022
% 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
% Ind. and Construction	0.054	0.027	0.014	0.056	0.090
% Transportation	0.029	0.016	0.004	0.032	0.048
% 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
% Non ess. - Service	0.032	0.013	0.016	0.032	0.047

3 Results

3.1 General Results

In this section, we present the main empirical results 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 as a proxy for the overall spread of infection within the neighborhoods. The first model includes some widely discussed potential factors of the spread of COVID-19 in NYC: density and commuting patterns, specifically, log of population density, percentage of workers using public transport, and average commute time. We also include the percentage of the population who is uninsured to control for those who do not have access to healthcare insurance. Our second model expands by including our proposed mechanism, namely, the percentage of the 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 of them, allowing us to detect any time patterns in the correlations. Therefore, in all of our specifications we run the following regression equation

$$\text{share of positive tests}_{it} = \alpha_t + \beta_t \text{tests per capita}_{it} + \gamma_t X_i + \epsilon_{it},$$

where the set of controls X_i vary according to the description below.

The first model shows the effect of the variables commonly used to explain the incidence of COVID-19 in NYC. Whereas Harries (2020) finds subway use was a major factor of the virus spread, we find it does not have a significant effect. This result could be due to the lack of cross-neighborhood variation to identify this effect, because most New Yorkers use public transportation in their daily commute. Nonetheless, commute time is a significant factor. For example, for April 20 a four-minute increase in commute time, a 10% increase on average, correlates with a 0.02-point increase in the share of positive tests, equivalently to approximately a 4-percentage-point increase in the share of positive tests.[11] We also find a positive and significant effect of the share of the uninsured population on the rate of positives for most of our sample. This result may be explained by uninsured patients only being willing to be tested under very acute symptoms in the fear of medical charges. For April 20, we find that under specification (1), a one-percentage-point increase in the share of uninsured population being correlated with a 1.7-percentage-point increase. Although the magnitude of this variable decreases as we include other covariates, its estimated coefficient remains positive and significant.

In specification (2) we test the importance of different occupations in the share of positive tests. We include the variables defined as the shares of the working-

[11] The average rate of positive tests on April 20 was 54%

age population employed in these occupations, so the coefficients are relative to the working-age but not employed population. The coefficients can be read as the effect of a one-percentage-point increase in the population employed in the particular category on the share of positive tests. We find some occupations explain a significant 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 the science fields category, legal occupations, and law enforcement have a negative correlation with the share of positive tests. These results are discussed further in the time-trends section.

Perhaps surprisingly, under this specification, commute time no longer has a significant effect. This result suggests commuting patterns are closely related to occupations, and most of the explanatory variation for commuting patterns may come 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 the share of positive tests and demographic characteristics, the results for specification (3) show that some of them can be explained through the occupation mechanism. Notably, the income effect disappears when we control for occupations, suggesting the previous correlation is due to income differences across jobs. Still, some demographic effects remain significant. These results are also robust after including borough fixed effects. For example, on April 20, a one-percentage-point increase in the share of Blacks and Hispanics leads to a 0.15% and 0.23% increase in the rate of positives, 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 positive correlation with test outcomes. On April 20, Adding one extra person to the average household, a 37% increase, corresponds with a 7% increase in the percentage of positive tests. Although neighborhood density does not explain variation in the share of positive tests, density in households appears to do so, with increasing magnitude over time.

The tests-per-capita coefficient is positive and highly significant across almost all specifications. Because of the scarcity of tests, testing was only available at earlier dates to those who required hospitalizations, and later on to those who were already showing symptoms. Therefore, we interpret this variable as a proxy for the rate of infections within the neighborhood, as the fraction of patients showing symptoms or requiring hospitalization is rather fixed for COVID-19.[12] Its magnitude decreases over

[12]A potential concern is large differences in age distribution across NYC zip codes. In the data, we find that the average age in NYC ranges from 27.5 to 45.5 across neighborhoods in NYC, with the exception of zip code 11005. It is a fairly small zip code with 1700 residents, an average age of 76, and mainly composed of retired immigrant women. Given such differences, we have excluded it

time as testing becomes more available and accessible to the rest of the population. We also find that its estimate remains significant even after including borough fixed effects.

from our analysis.

Table 3: Dependent variable - share of positive tests 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 ≥ 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 positive tests 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 positive tests 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 ≥ 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

In this section we present a time-variant analysis that could provide insights on both the evolution of the pandemic effects as well as the health policies in place. Figures 3 to 5 show the time evolution of the coefficients for specification (4). The result for the tests-per-capita variable is particularly salient; we observe a strong correlation 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, because it allows us to identify the mechanisms that explain demographic and occupational differences in COVID-19 exposure.

Notable time trends exist in the correlations associated with occupations. Higher shares of essential - professional and non-essential - service categories were associated with higher percentage-point increases in the rate of positive tests at earlier dates. On April 1, a one-percentage-point increase implied a 1.5- and a 2.2-percentage point increase in the positive rate of tests. However, they eventually decrease, averaging closer to a 0.3- and a one-percentage-point increase respectively on April 20, with essential - professional not being statistically significant. A plausible explanation is that these professions are either non-essential, or have the highest shares of remote workers. Although they were highly exposed to the virus in the beginning, once the workers shelter in place, their correlation with positive tests subsides. The opposite happens in science fields and law occupations — they are negatively correlated with COVID-19 incidence in the beginning, but the effect trends towards zero.

We find interesting patterns for the essential occupations as well. An additional percentage point in the share of transportation workers is associated with between a 0.5- and a one-percentage-point increase in the rate of positive tests. 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 correlation with COVID-19 incidence. However, a week after the general stay-at-home order, the governor determined construction was not essential, and this order could explain the eventual attenuation of the correlation. Law-enforcement-occupation shares have a consistently negative correlation on the share of positive, whereas firefighter shares have a declining trajectory toward zero. A plausible explanation for this difference could be the partnership between the NYPD and health care groups to provide free testing to its members.[13] Furthermore, the NYPD provided additional work flexibility for members with pre-existing conditions and extensive sick leave. It's possible that early adoption of these measures protected the most vulnerable workers from infection right from the onset. [14]

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

[14] More information on this can be read in www.policemag.com/548778/

The share of the uninsured population increasingly predicts the variation in positive test results. We find that an additional percentage point in the share of uninsured predicts an almost 0.3-percentage-point increase in the share of positive tests. Although many health care providers are waiving COVID-19-related out-of-pocket costs, these fees remain very high for the uninsured, and so a higher incidence of COVID-19 in this group could imply a severe financial burden. Although still significant, the effect of neighborhood density declines over time, 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 a stronger racial-selection component is at play 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.[15]

4 Conclusions and policy implications

In this paper, we present evidence showing that occupations are an important channel for explaining differences in the 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 to detect time patterns in their magnitudes.

We begin by showing descriptive evidence of the 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 positive tests. Conversely, we find that the shares of Black and Hispanic residents, and average household size positively correlate with the share of positive tests. Highlighting these differences is important because these observations confirm that the disease has had more harmful effects on vulnerable communities. Finding an occupation mechanism that explains it could guide policy measures intended to alleviate its impact.

We estimate several models to explore the effect of occupations. Our first specification only includes neighborhood characteristics, such as the use of public transportation and the average length of daily commutes. Although commuting patterns have been put forth as a 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 positive tests at the zip code level.

nypd-implements-policy-to-protect-most-vulnerable-officers-from-covid-19

[15] 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

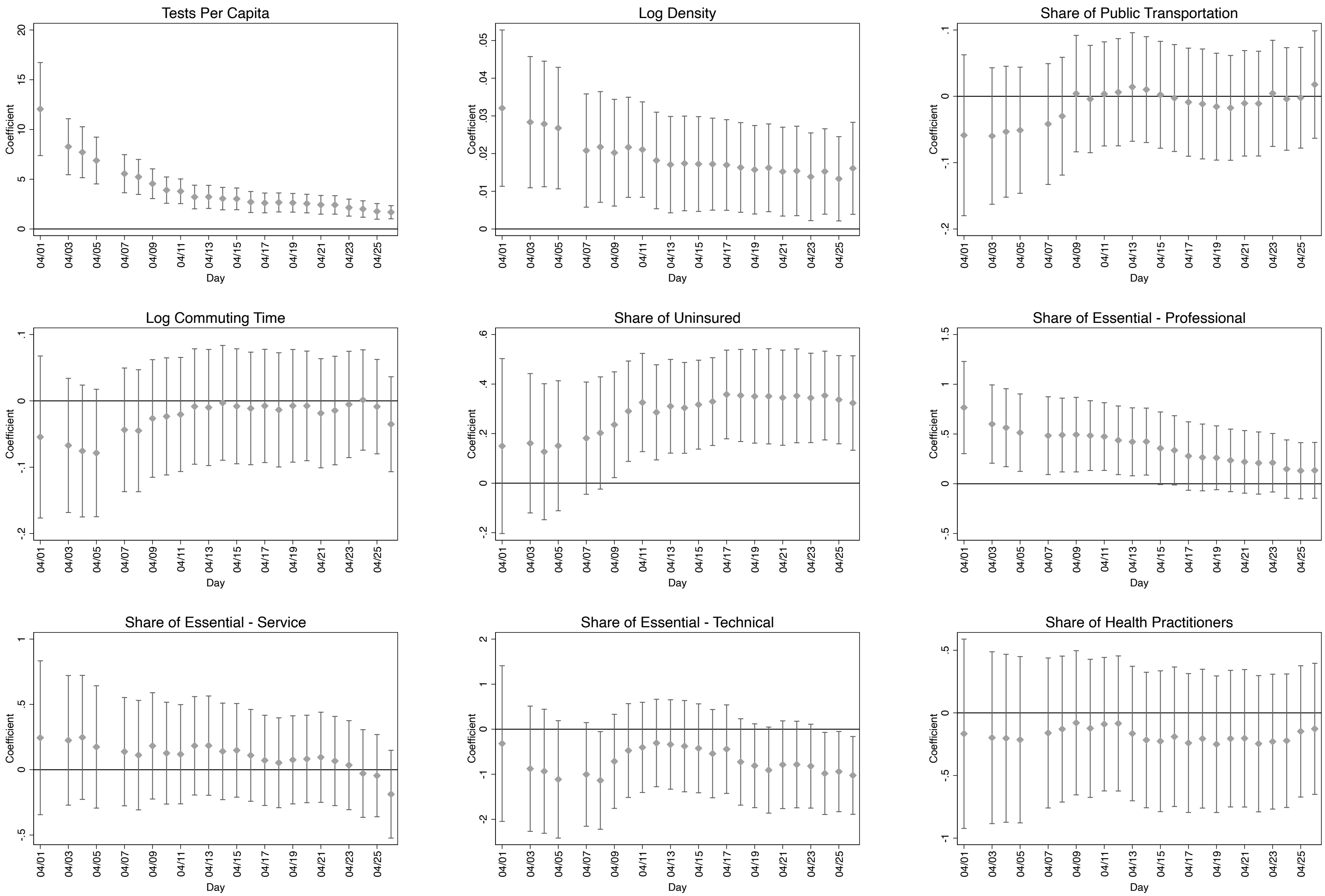

Figure 3: Regression coefficients of specification (4) over time

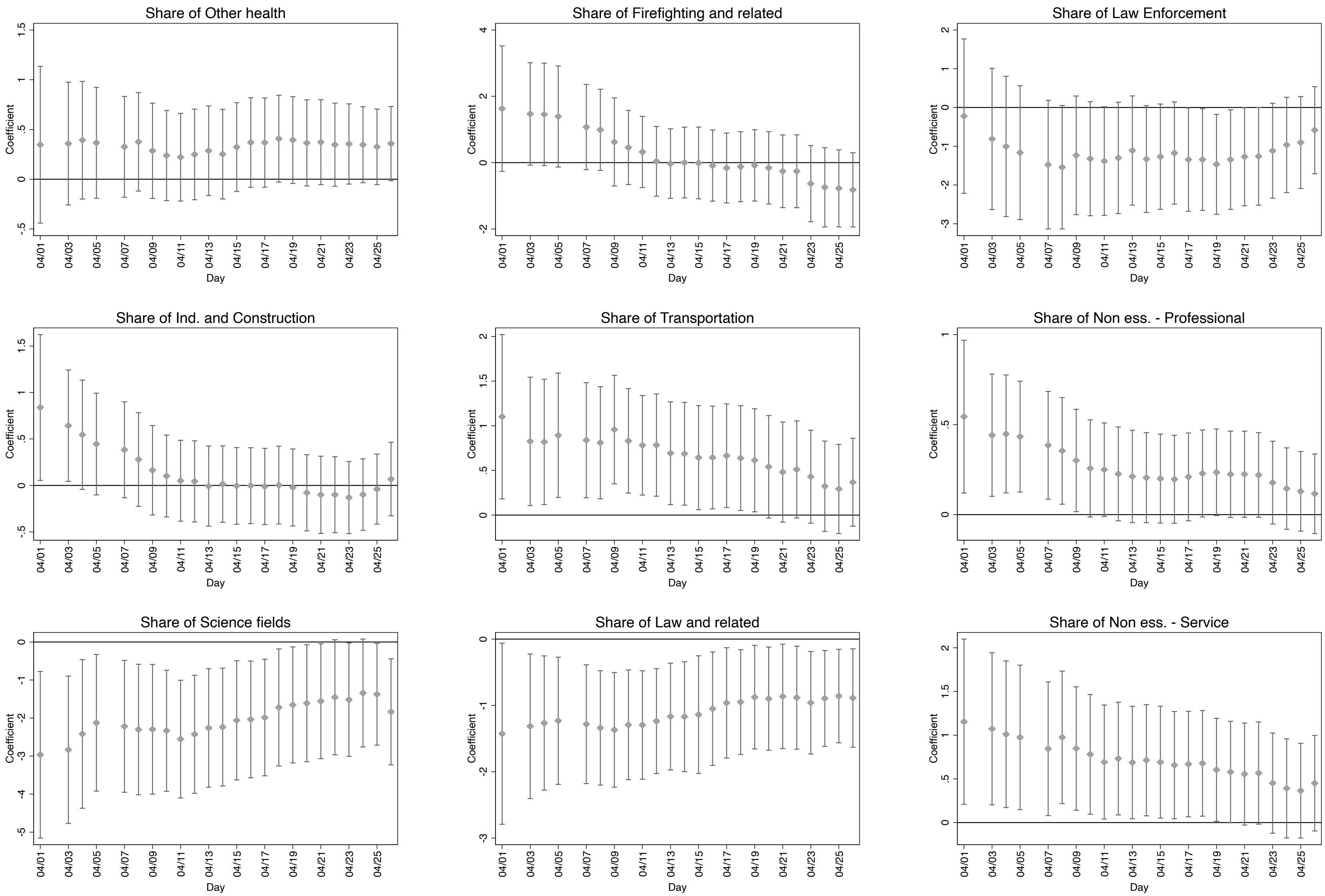

Figure 4: Regression coefficients of specification (4) over time

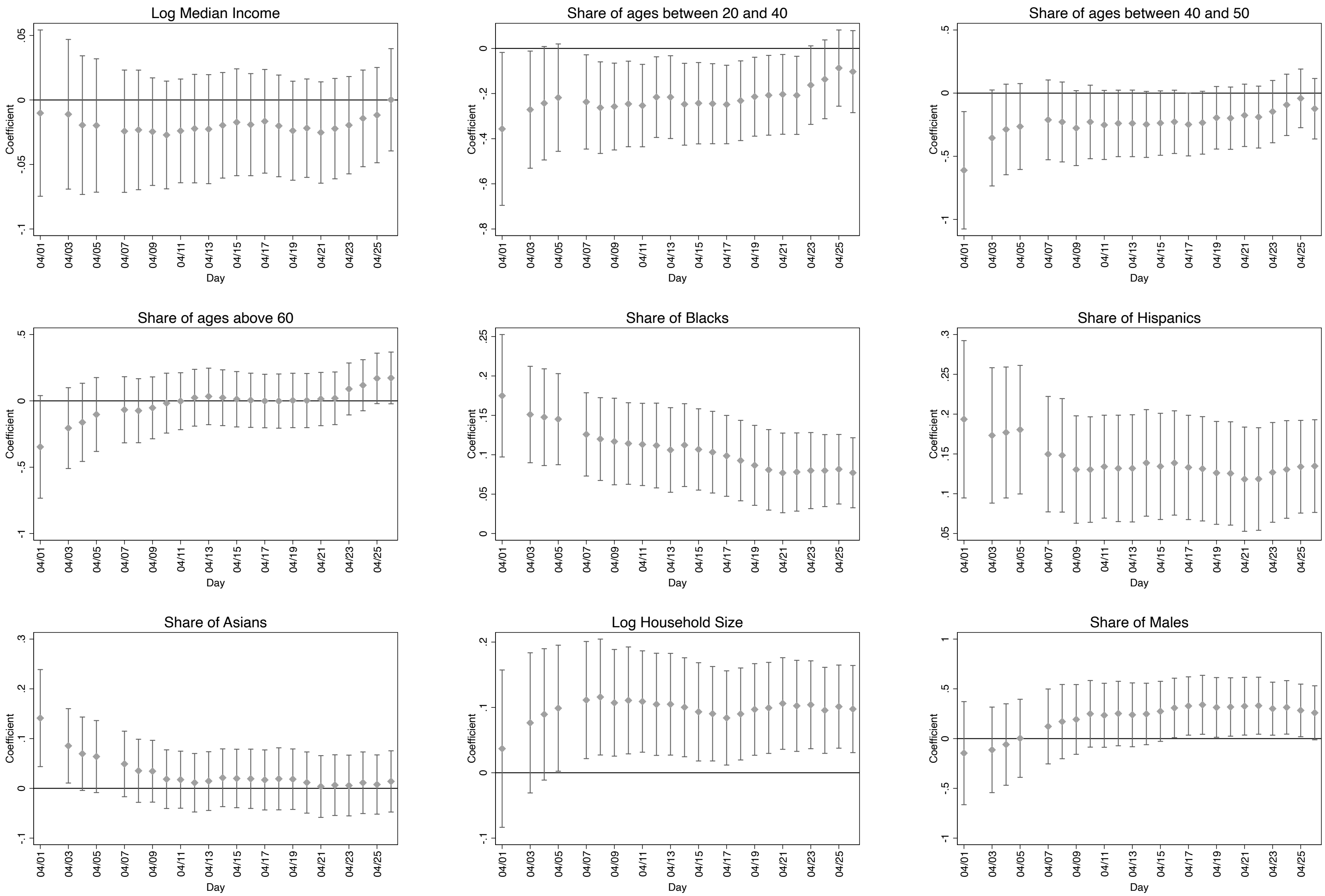

Figure 5: Regression coefficients of specification (4) over time

We find the strongest positive correlation on the share of positive tests with the share of workers in Transportation, Industrial, Natural-resources, Construction, and Non essential - Professional, with clear time trends in their estimated coefficients. For example, in the case of Transportation, a one-percentage-point increase in the share of workers in these occupations leads to a one- to two-percentage-point increase in the rates of positive results. Although the other two have a significant effect in positive shares at earlier dates, their magnitude becomes insignificant by the end of our sample period. This trend could be a result of the stay-at-home order. Conversely, higher shares of workers in Science Fields and Law Enforcement reduce the number of positive rates, with Science Fields decreasing in magnitude over time.

When adding demographic controls, we observe that racial patterns do persist, suggesting that the occupation mechanism does not fully explain all of the racial differences. However, their magnitude is small and arguably not economically relevant. Income and most age groups do not contribute to explaining the variation in positive tests, suggesting the occupation mechanism can explain to a greater extent the disparities along those demographics observed in the data.

In all of our regression models we include the number of tests per capita, and find that it is a strong predictor of the share of positive tests. However, its relative importance declines over time, as tests become more widely available. Moreover, as this variable loses relevancy, more of the variation in COVID-19 incidence is explained through the occupation channel.

Our results suggest clear implications for policy. First, they highlight the importance of mass testing in enabling clean identification of the relevant channels that increase the risk of infection. Second, once these channels are identified, policy-makers can target specific groups in the provision of protective gear, tests, and vaccinations. The purpose of this policy is twofold: while it provides extra protection against the disease for those who are more vulnerable, it also has positive spillovers on the rest of the population. For example, a policy that starts vaccinating and/or testing those workers with higher rates of human interaction affects not only those directly targeted by the policy, but also those who are likely to be in contact with them. Moreover, our results also suggest that health insurance condition, namely lack of insurance, plays a significant role, and its importance increases over time. Hence, local governments could incentivize the population without medical insurance to get tested, implementing policies such as full coverage of out-of-pocket costs in relation to COVID-19. Finally, we provide suggestive evidence that the stay-at-home order has mitigated contagion rates at work or in public spaces, while it has increased the probability of intra-household infections. This last result suggests the importance of policy or guidance measures to decrease spread within households.

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*.

Kaplan, G., Moll, B., and G., V. (2020). Pandemics according to hank. Available at `https://benjaminmoll.com/wp-content/uploads/2020/03/HANK_pandemic.pdf`.

Leibovici, F., Santacreu, A. M., and Famiglietti, M. (2020). Social distancing and contact-intensive occupations. Available at `https://www.stlouisfed.org/on-the-economy/2020/march/social-distancing-contact-intensive-occupations`.

Michaels, G., Rauch, F., and Redding, S. J. (2019). Task specialization in us cities from 1880 to 2000. *Journal of the European Economic Association*, 17(3):754–798.

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