Journal of Public Economics 194 (2021) 104328

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Divided we stay home: Social distancing and ethnic diversity ☆

Georgy Egorov [a], Ruben Enikolopov [b,*], Alexey Makarin [c], Maria Petrova [d]

[a] Northwestern University–Kellogg School of Management and NBER, United States
[b] New Economic School, ICREA-UPF, Barcelona IPEG, and Barcelona GSE, Spain
[c] Einaudi Institute for Economics and Finance (EIEF) and CEPR, Italy
[d] ICREA-UPF, Barcelona IPEG, Barcelona GSE, and New Economic School, Spain

ARTICLE INFO

Article history:
Received 15 August 2020
Revised 20 October 2020
Accepted 27 October 2020
Available online 23 January 2021

JEL classification:
D64
D74
I12

Keywords:
COVID-19
Pandemic
Social distancing
Self-isolation
Quarantine
Fractionalization
Diversity
Altruism
Xenophobia
Russia

ABSTRACT

Homogeneous societies usually provide more public goods. Voluntary social distancing in a pandemic is also a public good, but it has private benefits, too. Theoretically, we show that presence of population groups with different rationales for social distancing can lead to stricter observance of social distancing in more diverse societies. Empirically, we find that mobility reduction following the first local COVID-19 case was stronger in Russian cities with higher ethnic fractionalization and xenophobia. For identification, we predict the timing of the first case using historical patterns of internal migration. Using the United States data on mobility produces similar results.

1. Introduction

Prosocial behavior becomes commonplace in society either through government regulation or through voluntary adherence to social norms. Social distancing and self-isolation during a pandemic are two examples of such prosocial behavior; they play a key role in slowing the spread of infection. During the COVID-19 pandemic, governments in almost all affected countries have imposed restrictions aimed at promoting social distancing. However, enforcing these restrictions is logistically and politically costly. Thus, the effectiveness of these measures depends heavily on people voluntary observing social distancing guidelines. The conventional wisdom is that informal social norms are more difficult to sustain in ethnically diverse societies than in more homogeneous ones (Alesina and La Ferrara, 2000; Algan et al., 2016; Goette et al., 2006; Miguel and Gugerty, 2005; Putnam, 2007). This paper challenges that notion, by showing that during the COVID-19 pandemic prosocial behavior has increased *more* in ethnically diverse communities in Russia and the United States. Furthermore, we propose a theoretical mechanism to explain these findings.

Suppose that people belong to one of two ethnic groups and have one of three health statuses: they can be sick, healthy, or asymptomatic carriers. Sick people know they are sick and thus can't be infected, so their only reason to self-isolate is concern for others in the community. Healthy and asymptomatic carriers do not know whether they are infected, and their reasons to self-isolate are twofold. First, if they are healthy, self-isolation allows them to remain healthy. Second, if they are asymptomatic carriers

☆ We are grateful to Nadezhda Belyakova for excellent research assistance. We thank the Editor, two anonymous referees, Luigi Guiso, Timur Natkhov, Leonid Polishchuk, Michael Poyker, Andrey Simonov, and seminar participants at HSE Center for Institutional Studies for helpful comments.

* Corresponding author.
E-mail addresses: g-egorov@kellogg.northwestern.edu (G. Egorov), renikolopov@nes.ru (R. Enikolopov), alexey.makarin@eief.it (A. Makarin), maria.petrova@upf.edu (M. Petrova).

https://doi.org/10.1016/j.jpubeco.2020.104328
0047-2727/

who believe they can transmit the disease, they may have altruistic reasons to self-isolate. Notice that self-isolation decisions by healthy and sick people are strategic substitutes: if sick people are likely to self-isolate, healthy ones should feel freer to go out, whereas if healthy people stay at home, sick ones are less likely to infect anyone and should go out more. Suppose that asymptomatic transmission is being underestimated or dismissed. Then in a more diverse society, where sick individuals care less about others and are therefore less likely to self-isolate, healthy and asymptomatic individuals will be motivated by private benefits, which will induce them to self-isolate, given that the sick fail to do so. As long as most people are healthy, more ethnically diverse places should exhibit more compliance with self-isolation.

In this paper, we provide a theoretical model to formalize this argument and present causal evidence of the differential decline in social distancing by community's ethnic diversity in Russia and the United States. Our main empirical hypothesis is that, once an outbreak worsens and the threat of getting infected becomes real, people are more likely to minimize their day-to-day movements in places with higher ethnic diversity. To identify the effect, we rely on a discontinuous jump in the perceived threat of getting infected after the first case of COVID-19 was reported nearby.[1] Moreover, as Russia is the world's largest country by territory, there was substantial heterogeneity in the timing of the first case, from March 2 in Moscow to April 16 in Altai Republic. Thus, our empirical approach allows us to include day fixed effects and differentiate between the effect of Covid-19 and common shocks.

However, the timing of the official reporting of the first local case is potentially endogenous. It may be affected by the quality of the medical system (e.g., its capacity to diagnose) or by officials' willingness to publicly admit the problem,[2] both of which can affect citizens' willingness to observe social distancing guidelines.[3] To deal with this potential endogeneity problem, we use the fact that preexisting internal migration patterns predict travel flows in 2020. Therefore, how soon the virus spreads to different locations can be predicted by internal migration (Mikhailova and Valsecchi, 2020; Valsecchi and Durante, 2020). In Russia, the coronavirus spread primarily from Moscow, which methodologically allows us to use two-stage least squares. To predict the timing of the first case, we relate it to internal migration flows to Moscow. Following the literature on migration in labor economics (e.g., Altonji and Card, 1991; Card, 2001), we use a shift-share instrument for internal migration. In particular, we combine the data on migration from a given region to Moscow during the 1990s with the nationwide domestic out-migration from that region in more recent years (2015–2018) to instrument for the recent migration flows to Moscow. We then predict the timing of the first local case using the instrumented migration flows from a given region to Moscow. Finally, we use the predicted timing of the first coronavirus case in the region in a difference-in-differences framework, comparing people's behavior before and after the predicted discovery of the first case in places with different levels of ethnic diversity. Note that internal migration to other large cities does not significantly explain the timing of the first COVID-19 case in a region, consistent with the disproportionately high presence of the virus in Moscow.[4]

We use data on people's movements provided by Russia's largest technology company, Yandex, which tracks individuals' cell phones with its mobile apps. We find that people have been more likely to engage in social distancing since the first local COVID-19 case report in more ethnically diverse places. Numerically, we find that a one-standard-deviation increase in ethnic fractionalization explains 5.7% of mobility reduction after the first case report. This magnitude corresponds to 4.7% of the average weekday-weekend gap in mobility. Importantly, these magnitudes change little after controlling for government-imposed mobility restrictions.

To provide more evidence on the mechanisms behind our findings, we test whether ethnic intolerance produces an additional distancing effect beyond ethnic diversity. To measure ethnic tensions, we use data on xenophobic online searches and the number of ethnic hate crimes in a city in recent years. The results confirm that the reduction in mobility after the first reported case is stronger in places with more xenophobic searches, as well as in places with more hate crimes, even after accounting for ethnic fractionalization. We then show that some alternative stories, e.g., differences in social capital or quality of the medical system, cannot explain our results.

We ensure that our findings are not specific to Russia by obtaining very similar results for differential mobility reduction in the United States.

Finally, to put our estimates in perspective, we produce a back-of-the-envelope calculation of how many premature deaths might have been prevented by a stronger social distancing response in more diverse communities. We rely on two estimates of the effect of social distancing on the eventual number of COVID-19 deaths: one from a mainstream epidemiological model by Walker et al. (2020) and one from the local average treatment effect estimated by Kapoor et al. (2020), based on a rainfall IV strategy. We consider the elasticity produced by Walker et al. (2020) as the upper bound, because they take into account all potential future deaths from the disease that evolve according to their model. In contrast, we consider the elasticities in Kapoor et al. (2020) as the absolute lower bound, because they study a temporary reduction in social distancing on one particular weekend and because they rely only on data available at the time they wrote their article. Based on these two studies, we calculate that a one-standard-deviation increase in ethnic fractionalization is associated with 570 to 22,250 fewer deaths in Russia and 2,000 to 40,000 fewer deaths in the United States.

Our paper contributes to the literature on the role of voluntary adherence to social norms in establishing order in society (Ostrom, 1990; Ellickson, 1994). Cooperation based on other-regarding preferences helps sustain informal institutions and social norms, which greatly enhance the possibilities of collective action (Fehr and Gächter, 2000). The existing literature suggests that the informal social norms are more difficult to maintain in ethnically diverse societies (Alesina and La Ferrara, 2000; Miguel and Gugerty, 2005; Goette et al., 2006; Putnam, 2007; Algan et al., 2016). Our paper shows that, contrary to this conventional wisdom, voluntary social distancing during the pandemic may be higher in more diverse places, due to the coexistence of public and private benefits from the prosocial action.

We also contribute to the literature on the impact of diversity on development outcomes. Ethnic diversity is often found to have deleterious effects on economic growth (Easterly and Levine, 1997; Alesina and Ferrara, 2005), public good provision (Alesina et al., 1999), and civil conflicts (Montalvo and Reynal-Querol, 2005; Rohner et al., 2013; Arbatli et al., 2020).[5] However, evidence has emerged in recent years showing that diversity can also be beneficial

[1] Barrios and Hochberg (2020), for example. document a significant increase in COVID-19 Google searches on the day of the announcement of the state's first case in the United States. Campante et al. (2020) find a similar pattern during the 2014 Ebola scare in the United States.

[2] Note that there is a general skepticism about the accuracy of Russian coronavirus statistics, both in terms of deaths and cases (see, e.g., Dixon (Washington Post, 29/06/2020), Roth (Guardian, 04/06/2020), or Gabow (Deutsche Welle, 21/06/2020)).

[3] With few exceptions, the virus hit more densely populated and economically developed places first; see, e.g., Desmet and Wacziarg (2020).

[4] Moscow accounted more than 50% of all reported cases in Russia in the time period that we study (see Fig. B.6).

[5] Intergroup tensions induced by conflict are also found to decrease interethnic team performance (Hjort, 2014) and harm intergroup trade (Korovkin and Makarin, 2019).

for productivity (Ottaviano and Peri, 2006; Peri, 2012), innovation (Lee, 2015), and economic development (Alesina et al., 2016, 2018, 2020). When governments fail to reach their policy goals, they often blame ethnic cleavages. Our paper shows that group heterogeneity can help governments reach their policy goal of imposing social distancing, through better individual adherence to this behavior.

Finally, we contribute to the emerging literature on the determinants of social distancing and compliance with stay-at-home orders during the COVID-19 pandemic. This literature is mostly based on a difference-in-differences analysis, with social distancing as a function of stay-at-home orders and some third variable. For instance, several studies focusing on the United States independently find that Republican-leaning counties comply less with social distancing recommendations and quarantine orders (Allcott et al., 2020; Andersen, 2020; Barrios and Hochberg, 2020; Engle et al., 2020; Painter and Qiu, 2020; Wright et al., 2020).[6] Other factors that predict lower compliance with social distancing measures include lower local infection rates and younger populations (Engle et al., 2020), poverty (Wright et al., 2020), inequality (Gitmez et al., 2020) as well as mistrust in science and lower education levels (Brzezinski et al., 2020).[7] However, given that both stay-at-home measures and coronavirus spread are unlikely to occur at random, identification remains an important concern. Both the spread of the virus and the slate of emerging policies will likely depend on the dynamics of healthcare system capacity and testing protocols. We improve on this literature by explicitly addressing the issues of the endogeneity of the timing of the spread of the virus.

The rest of the paper is organized as follows. Section 2 contains some background information about Russia and its response to COVID-19. Section 3 summarizes theoretical results. Section 4 presents our empirical strategy and data. We present our main results in Section 5 and additional results in Section 6. Section 7 concludes. The not-for-publication Appendix consists of four parts: Appendix A provides additional information on the data sources, Appendix B presents additional evidence and illustrations, Appendix C presents the complete model and proofs of all results, Appendix D replicates our main results with the U.S. data, and Appendix E discusses the implications of our results for mortality from COVID-19.

2. Background

The first cases of novel coronavirus-infected pneumonia, now known as COVID-19, were detected in China's Hubei Province in December 2019. Within weeks, the virus had spread throughout East Asia, then jumped to the United States. The infection arrived to Russia from Italy in early March (the few earlier cases of arriving passengers from China and tourists from the Diamond Princess had been quickly isolated). During the period we analyze the United States and Russia had the two largest numbers of reported cases.

Moscow quickly became the epicenter of the pandemic in Russia, other Russian regions typically got the disease from people arriving from Moscow. Despite abundant international news and evidence, Russian citizens were generally skeptical of the coronavirus threat and distrusted information on it put out by the media or the government. Discovery of regional COVID-19 cases played a much bigger role in informing local populations of the reality and severity of the virus.[8]

Although Russia is commonly thought of as ethnically homogeneous, it is a multinational country that is home to dozens of ethnic minorities with plenty of regional ethnic heterogeneity. According to the 2010 Census, ethnic minorities comprise 19.1% of the Russian population. On one end of the spectrum are the nearly homogeneous Yaroslavl and Novgorod oblasts, with 96% and 93% of Russians, respectively; meanwhile, the Republic of Tatarstan is ethnically highly heterogeneous, with 115 ethnicities, a Tatar majority (53.2%), and a sizeable (39.7%) Russian minority.

3. Model

Society consists of two ethnic groups, G_1 (share $g \in (0, \frac{1}{2}]$) and G_2 (share $1-g$). In the beginning, each individual is either healthy (share h), sick (share s) or an asymptomatic carrier (share c); the health status is assumed to be independent of ethnicity. Individuals observe whether they are sick, but not if they are healthy or an asymptomatic carrier.

Each individual makes a binary decision $d_i \in \{0, 1\}$, where 1 stands for self-isolation and 0 for going out. Going out has a direct benefit b_i, distributed on uniformly on $[0, W]$ (we can allow the distribution to be different for sick and healthy/asymptomatic carriers). The cost of going out depends on one's health status. A healthy individual may become infected, and anyone who is infected at the end of the game gets utility $-L$. An infected person (sick or asymptomatic) might infect someone else, leading to a psychological cost M per each healthy person infected from the same ethnic group; the cost of infecting an out-group person is tM, where $t \in [0, 1]$ captures tolerance toward the other ethnic group.

Consider the following simplified model of interactions during a pandemic. Suppose that all individuals are matched in pairs, and let $m(i)$ denote the match of individual i. Assume that if everyone goes out, then each i would come in close contact with exactly one other person, their match $m(i)$. If one or both of two matched individuals decide to stay home, there would be no transmission of the infection between them. The same is true if they both go out, if both are healthy or both are infected (regardless if they are carriers or are sick). If one is healthy and the other is infected, the healthy one becomes infected with probability q if the infected person is sick and r if the infected person is a carrier.[9] Naturally, $r > 0$ reflects the possibility of asymptomatic transmission.

We assume that $qM < W < qL$. This is satisfied if the disutility of getting infected is high enough, so nobody would go out if they were certain to encounter a sick individual (and get infected with probability q), and if the psychological cost is not too big, guaranteeing that at list some sick people will go out.

In Appendix C, we spell out the maximization problems of all individuals and formally show that there is a unique perfect Bayesian equilibrium. Our main result is the following.

[6] The emerging literature also studies the impact of persuasion on people's mobility. For example, Simonov et al. (2020) and Ananyev et al. (2020) independently show that higher Fox News viewership has led to a significantly lower propensity to stay at home during the pandemic. Bursztyn et al. (2020) show that, even conditional on viewing Fox News, watching TV hosts who are more concerned about COVID-19 (e.g., Tucker Carlson) has led to fewer coronavirus cases and deaths.

[7] Interestingly, the evidence of the effect of social capital and trust on voluntary social distancing is mixed. Borgonovi and Andrieu (2020), Barrios et al. (2020), and Durante et al. (2020) document evidence of a larger drop in social mobility in areas with higher social capital in the United States and Italy, but Doganoglu and Ozdenoren (2020) provide cross-country evidence that generalized trust is associated with *less* social distancing.

[8] A survey revealed that 60% of Russians trust information about the coronavirus from doctors they know personally, while only 8% trust information from the Russian Ministry of Health (https://www.rbc.ru/society/17/04/2020/5e998b669a794768d09da79e).

[9] The probability of getting infected is thus proportional to the mass of infected individuals who go out, weighted by their contagiousness. In practice, this relationship may be more complex. For example, it may be concave because of the possibility of getting infected by multiple individuals, or it might be convex, for example, because close interactions are easier to avoid when few sick people are out. We adopt the simple proportionality assumption for simplicity.

Proposition 1. *If* $W \leqslant q\frac{h}{h+c}sL$, *then all people without symptoms self-isolate, and all sick people go out. Otherwise, the equilibrium is interior, with positive shares of all types self-isolating and going out. In this case, an increase in the size of the minority group g, a decrease in altruism M, or a decrease in tolerance t all decrease self-isolation by sick individuals. The effect of these changes on overall self-isolation is positive if* $\frac{rc}{qs} < \frac{qhL-W}{W+q\frac{h}{h+c}sL}$, *and negative otherwise.*

The right-hand side of the last condition is positive for h close to 1, i.e., in the beginning of the pandemic. This means that the comparative statics critically depends on the likelihood of asymptomatic transmission (and the share of asymptomatic people) relative to the likelihood of transmission from sick people (and their share). If this ratio is small, then higher fractionalization implies less self-isolation by sick individuals, but more self-isolation overall, because healthy individuals are concerned about getting infected by sick people who self-isolate less. If, however, asymptomatic transmission is a major issue, then higher fractionalization also means that people without symptoms are less concerned about infecting healthy ones, and thus overall self-isolation may decrease. As h becomes small (e.g., later in the pandemic), the comparative statics becomes driven solely by sick individuals, and fractionalization will imply less self-isolation. The effect of a decrease in altruism or tolerance is similar.[10]

4. Empirical strategy

Our theory predicts that people engage in social distancing more in places with higher ethnic fractionalization if the likelihood of asymptomatic transmission is relatively low or when the probability of getting infected becomes nontrivial. We test this prediction empirically. First, we report difference-in-differences estimates, where we compare cities with higher and lower levels of ethnic fractionalization before and after the first reported case of COVID-19 infection in their region. Second, we combine the difference-in-differences approach with a two-stage least-squares approach, in which the timing of the first reported case is instrumented using measures of preexisting migration.

We aim to estimate the following specification:

$$SocialDistance_{irt} = \alpha_i + \theta_t + \gamma FirstCase_{rt} + \beta FirstCase_{rt} \times Ethnic_i + \mathbf{X}_{irt}\delta + \epsilon_{irt}. \quad (1)$$

Here, $SocialDistance_{irt}$ is a measure of people's staying at home (or lack of mobility) in locality i in region r at time t. This measure is based on daily averages of the Yandex Isolation Index, which aggregates data on people's movements at the city level and which is analogous to the Google Mobility Index. The index is calibrated for each city to be 0 for the busiest hour of the working day, and 5 for the quietest hour of the night before the coronavirus outbreak.[11] We use daily data for 302 cities with a population over 50,000 from February 23, 2020 through April 21, 2020.

$FirstCase_{rt}$ is an indicator variable equal to 1 after the first reported case of COVID-19 in region r (first predicted case in the region in the 2SLS estimation). Recorded cases are taken from the government-agency website that contains official information about the pandemic. As shown in Fig. B.5 there was a significant variation in the timing of the first case in our data, which ranges from March 1 until April 16.

$Ethnic_i$ is an ethnic fractionalization index in locality i which measures the probability that two randomly drawn individuals in society belong to different ethnic groups. As such, the index varies between 0 and 1, with a mean of 0.22 and a standard deviation of 0.18 for the Russian data. X_{irt} is a vector of controls, which includes interactions of $FirstCase_{rt}$ with baseline locality characteristics; α_i are the locality fixed effects, which control for any time-invariant locality characteristics, such as population, population density, and baseline levels of income and education; and θ_t are the day fixed effects, which account for countrywide shocks. To adjust for regional correlations in the error term, ϵ_{irt}, standard errors are clustered at the level of Russian regions. Our sample of cities spans 83 Russian regions.

In the OLS specifications, we estimate Eq. (1) using the data on the dates of the first case. The identifying assumption is that of parallel trends—in the absence of the novel coronavirus, social distancing patterns in places with high and low ethnic diversity would have followed parallel trends. One potential concern with this approach is that the timing of the first case is not fully random. For example, regions could report late COVID-19 cases because their medical capacity precluded them from correctly identifying the virus in time, or because their testing policies could be ineffective, or because their administration was prone to conceal the first cases for longer. To deal with these potential confounds, we predict the timing of the first case in Eq. (1) using a two-stage least-squares framework.

Specifically, we use the fact that travel connections between various cities and Moscow (where the first major outbreak occurred) could affect the timing of the first case in those cities' respective regions. We rely on internal migration as a proxy for these type of connections (Mikhailova and Valsecchi, 2020; Valsecchi and Durante, 2020). We then estimate the following regression specification for the timing of the first case at the regional level:[12]

$$FirstCase_r = \alpha_0 + \alpha_1 MigToMoscow_r + \eta_r. \quad (2)$$

Here $MigToMoscow_r$ stands for recent migration flows from region r to Moscow, while $FirstCase_r$ is the date of the first case in this region.[13]

Next, we predict the timing of the first case from Eq. (2), create a dummy that is equal to 1 after the date of the predicted first case, and finally plug this variable into the Eq. (1) to estimate the second stage. To consistently estimate Eq. (2), we follow the migration literature (e.g., Altonji and Card, 1991; Card, 2001), creating a shift-share instrument for internal cross-regional migration. Specifically, we compute the following term:

$$\frac{EarlyMigrationToMoscow_r}{\sum_i EarlyMigrationToRegion_{ir}} \times RecentTotalMigationFromRegion_r, \quad (3)$$

then use it to predict $MigToMoscow_r$ in Eq. (2). Since this is not a standard IV procedure, for the second-stage estimation, which combines IV with difference-in-differences, we use the bootstrap

[10] Note that our model is also able explain seemingly contradictory evidence about the role of social capital in social distancing from Barrios et al. (2020), Borgonovi and Andrieu (2020), and Durante et al. (2020) as opposed to Doganoglu and Ozdenoren (2020). First of all, if asymptomatic transmission is not perceived to be likely, higher altruism makes sick people more likely to stay at home, while social capital could also imply higher benefits W from going out. Second, if asymptomatic transmission is perceived to be a large risk, everything is driven by the behavior of asymptomatic people, and higher altruism increases rather than decreases social distancing. Thus, for different parameter values, our model can rationalize both positive and negative effects of social capital on social distancing found in the literature.

[11] Fig. B.1 demonstrates the change in isolation for Moscow between February 23 and May 5, 2020.

[12] Note that we only have dates for the first case and data on internal migration flows only at the regional rather than the city level.

[13] Both Moscow and Saint Petersburg have regional status in the Russian administrative division, in contrast to most other cities, which are administratively parts of their region. Thus region-level statistics on internal migration includes the data on migration to Moscow.

method to compute standard errors. The identifying assumption behind this strategy is that the migration to Moscow from a particular region during the 1990s—interacted with recent (2015–2018) total outflow of migration for this region and further interacted with ethnic fractionalization in a city—affects isolation only through the timing of the first case interacted with ethnic fractionalization (conditional on city and day fixed effects).

5. Empirical results

Parallel Trends. Identification in the OLS estimation of Eq. (1) relies on the parallel trends assumption, which implies that in the absence of COVID-19, people's patterns of movement and of staying at home would evolve in parallel fashion for places with different levels of ethnic fractionalization. This assumption is not testable, but we can provide some supportive evidence by examining pretrends. Fig. 1 summarizes the patterns of people's movements before and after the first case in a region, conditional on city and day-of-the-week fixed effects.

Fig. 1 shows no visible difference in the behavior of people in the two groups of cities before the first coronavirus case. In both groups of cities, people have engaged in more social distancing since the discovery of the first case. However, we detect a marked difference: after one week people in more fractionalized cities have been more likely to stay home than people in less fractionalized cities. These results are consistent with the parallel trends identifying assumption for (1). The effect does not manifest itself immediately after the discovery of the first case, which likely reflects the fact that a certain time period is needed to disseminate information about the discovery of the coronavirus in the region. Moreover, the growth in self-isolation in more fractionalized cities is somewhat lower in the first days after the discovery of the first case, which may be driven by people catching up on unfinished tasks that require mobility, such as last-minute purchases, in anticipation of more stringent self-isolation in the future. Overall, this preliminary evidence already favors our main empirical hypothesis.

Baseline difference-in-differences results. Here, we report the results of estimation of Eq. (1) using ordinary least squares. Table 1 summarizes these results. Column 1 reports the basic specification with city fixed effects, day-of-the-week fixed effects, and calendar-week fixed effects included. Column 2 adds several additional controls, specifically the interactions of the *Post First Case* dummy with shares of people with higher education, average wage, and population density. Columns 3–4 report the same specifications with day fixed effects included instead of day-of-the-week and calendar-week fixed effects. The results indicate that the coefficient for the interaction between the *Post First Case* dummy and ethnic fractionalization is consistently positive and significant in all the specifications. The magnitude of the coefficient goes slightly down from 0.38 to 0.32 with additional interactions, but it remains statistically significant at the 1% level. This reduction is smaller than the standard error for both coefficients, and we cannot reject the hypothesis of the equality of the coefficients in a seemingly unrelated regressions framework. Thus, we conclude that the coefficient is robust to inclusion of additional controls. The results in Table 1 are consistent with our theoretical prediction: we observe more social distancing in more ethnically diverse places. The magnitudes in Table 1 imply that a one-standard-deviation increase in ethnic fractionalization [14] leads to 4.2% higher social distancing following the report of the first local COVID-19 case. In other words, a one-standard-deviation increase in ethnic fractionalization can explain 6.5% of the average mobility reduction after the report of the first case or, alternatively, 5.3% of the weekday-weekend gap for an average locality.

IV estimation. First stage. As discussed above, the OLS estimates from the previous subsection could be biased because of the endogeneity of reporting of the first case in a region, which would lead to the violation of the parallel trends assumption. In what follows, we estimate Eq. (1) using the 2SLS approach. We first check whether our logic for the first stage holds, and internal migration to Moscow indeed predicts the timing of the first case in the region. In particular, we estimate Eq. (2) using OLS and IV, using the shift-share instrument (3) to predict migration in the latter case.

The results of these estimations are summarized in Table 2. Columns 1–2 present the results of the OLS estimation, and columns 3–4 present the results of the IV estimation. Columns 1 and 3 present the results without additional controls, while columns 2 and 4 contain the results with basic controls such as population density, income, and education. The results suggest that migration to Moscow has a large negative effect on the timing of the first case. The coefficient is remarkably stable when extra controls are added. IV coefficients are slightly larger than OLS ones, with the magnitudes of coefficients going from -58.93 for OLS to -66.80 for IV. These magnitudes imply that a one-standard-deviation increase of internal migration to Moscow led to the first case being reported 4.6 days earlier according to the OLS estimates, or 5.2 days earlier according to the IV estimates. Another important predictor of the timing of the first reported COVID-19 case is the average wage. According to the estimates, a one-standard-deviation increase of average wage led to the first case 2.5 days earlier.

We also check whether migration to Moscow played a special role in spread of the virus across the country compared with migration to other big cities. As Figs. B.6 and B.7 suggest, Moscow accounted for the disproportionately large share of all COVID-19 cases, compared to its share of internal migration (as well as its share of the country's population, which is around 10%). This suggests that while other large cities could play a similar role, their importance in spreading the virus was likely smaller. In Table B.3, we report the results of this estimation. The coefficients for migration to the regions with other large cities are smaller in magnitude and flip signs if additional controls are added. Neither the OLS nor the IV coefficients are significant in this estimation, although our shift-share instrument still works reasonably well, with the corresponding Kleibergen-Paap F-statistic around 200 (see columns 3–4). Similarly, Fig. B.8 in the Appendix reports the distribution of the first-stage coefficients for each Russian city with a population above 1 million. Outmigration to Moscow indeed plays a special role, with the first-stage coefficient being negative, highly significant, and three times larger in magnitude than any other coefficient for the other cities, while the rest of the coefficients are small, mostly insignificant, and are positive or negative with approximately the same probability.[15] The results of Table B.3 and Fig. B.8 confirm the special role of Moscow and regional links to Moscow in the spread of the virus, consistent with the idea that tighter migration connections to Moscow resulted in regions getting the coronavirus earlier.

2SLS estimation. Second stage. Once we predict the timing of the first case, as summarized in columns 3–4 in Table 2, we can now use these predicted values in the second-stage estimation.

[14] One standard deviation ethnic fractionalization equals 0.18, which is smaller than the mean ethnic fractionalization in the sample (0.22) and higher than its median (0.14). A one-standard-deviation increase in ethnic fractionalization is equivalent to moving from a fully homogeneous city to a city with two groups where a minority group comprises 20 percent of the population.

[15] In Russian administrative division, Moscow and St Petersburg have a status of regions, while all other cities are parts of their corresponding regions. As a result, the reported numbers are for migration to Moscow and St Petersburg for those two cities or for migration to corresponding regions for the other cities.

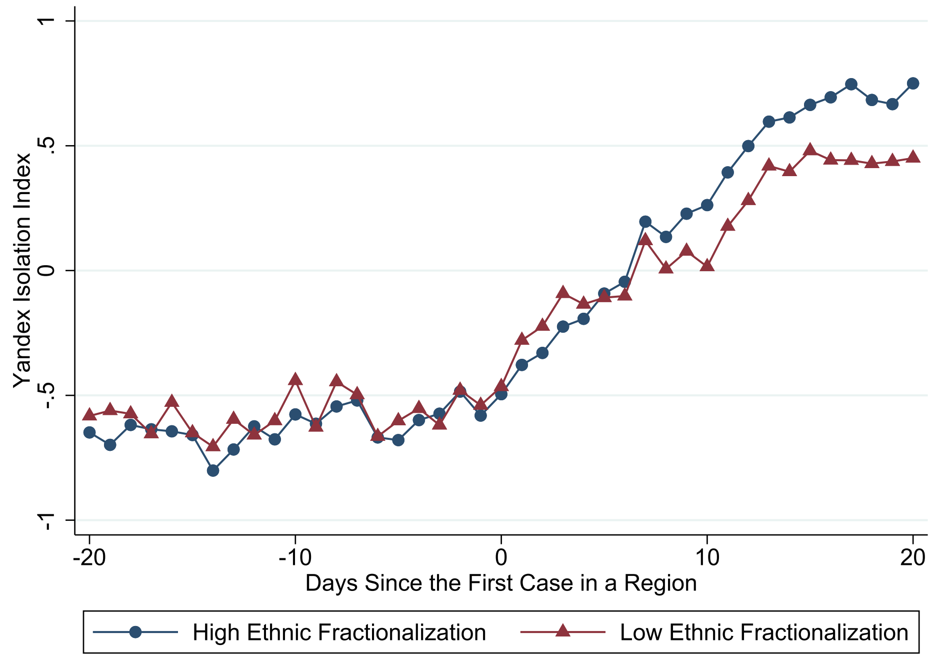

Fig. 1. Isolation Over Time for Places with High and Low Ethnic Fractionalization, Russian Data. *Notes*: The Yandex isolation index is de-meaned by city and day-of-the-week fixed effects. *Source:* Authors' calculations.

Table 1
Social Distancing, First Case, and Ethnic Fractionalization (OLS).

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	0.378***	0.318***	0.380***	0.324***
	[0.111]	[0.078]	[0.113]	[0.091]
Post First Case	−0.037	1.233**	−0.095*	0.808
	[0.068]	[0.515]	[0.050]	[0.593]
Post First Case x Education		1.880***		1.818***
		[0.263]		[0.266]
Post First Case x Average Wage		−0.180***		−0.142**
		[0.055]		[0.063]
Post First Case x Population Density		0.003**		0.003**
		[0.001]		[0.001]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,817	17,817	17,817	17,817
R-squared	0.816	0.820	0.944	0.948

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in brackets are clustered by region. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020.

We report the results of this estimation in Table 3 below. The results of the 2SLS estimation are similar to the results of the OLS estimation, with more social distancing post-outbreak in cities with higher ethnic fractionalization. The 2SLS magnitudes are slightly smaller than the OLS ones, but we cannot reject the hypothesis of the equality of the coefficients. The magnitudes in Table 3 imply that a one-standard-deviation increase in ethnic fractionalization leads to 3.7% higher social distancing following the report of the first local COVID-19 case. In other words, a one-standard-deviation increase in ethnic fractionalization can explain 5.7% of the average mobility reduction after the report of the first case or, alternatively, 4.7% of the weekday-weekend gap for an average locality.

Overall, the results in Tables 1 and 3 are consistent with the main theoretical prediction that higher ethnic fractionalization increases social distancing once the threat of the virus becomes real.

6. Additional results and mechanisms

Xenophobia. Our theory suggests that a reduction in tolerance towards out-group members should lead to more self-isolation, even holding the preexisting levels of ethnic diversity fixed.[16] We test this prediction using two distinct measures of xenophobia in Russian cities, one based on explicitly xenophobic Internet searches and the other based on the number of ethnic hate crimes in the earlier period. The results of these estimations, summarized in Table 4, indicate that both xenophobia and the history of ethnic hate crime led to an increase in social distancing following the discovery of the first COVID-19 case in the region. Moreover, both of these effects coexist with the positive effect of ethnic fractionalization without

[16] This channel's importance is further stressed by the recent evidence suggesting that the COVID-19 crisis has increased hostility toward foreigners (Bartos et al., 2020).

Table 2
Timing of First Case and Internal Migration to Moscow, 2015–2018.

	Date of the First Covid-19 case in a Region			
	OLS		IV	
VARIABLES	(1)	(2)	(3)	(4)
Migration to Moscow in 2015–2018	−59.676***	−58.934***	−68.697***	−66.979***
	[11.314]	[9.176]	[5.869]	[5.805]
Average Wage		−6.106**		−5.957**
		[2.451]		[2.416]
Education		15.645		17.042*
		[9.623]		[9.634]
Population Denisty		−0.018		0.011
		[0.043]		[0.047]
Observations	302	302	302	302
R-squared	0.372	0.410	0.364	0.406
Kleibergen-Paap F-statistic			2,032	4,102

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Robust standard errors in brackets are clustered by region. The sample includes 302 Russian cities with a population of at least 50,000. In columns (3) and (4), migration to Moscow is predicted with a shift-share instrument, using pre-1998 migration to Moscow combined with 2015–2018 aggregate outflow of internal migration from a region.

Table 3
Social Distancing, First Case, and Ethnic Fractionalization (2SLS).

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post Predicted First Case x Ethnic Fractionalization	0.352***	0.293**	0.345***	0.285**
	[0.109]	[0.122]	[0.107]	[0.117]
Post Predicted First Case	−0.154**	0.893	−0.186***	0.793
	[0.069]	[0.547]	[0.065]	[0.540]
Post Predicted First Case x Education		1.798***		1.813***
		[0.288]		[0.289]
Post Predicted First Case x Average Wage		−0.156***		−0.151***
		[0.059]		[0.058]
Post Predicted First Case x Population Density		0.003**		0.003**
		[0.001]		[0.001]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,817	17,817	17,817	17,817
R-squared	0.816	0.820	0.944	0.949

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Bootstrapped robust standard errors in brackets are clustered by region. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020. Predicted First Case is computed using the data on inter-regional migration, as summarized above.

canceling each other. The coefficients for xenophobic searches (Panel A of Table 4) and and ethnic hate crime (Panel B of Table 4) go down substantially when additional interaction terms with control variables are included, but the main coefficient for ethnic fractionalization remains remarkably stable in terms of its magnitude.

Stay-at-home orders. The results in Tables 1 and 3 could reflect that, following the coronavirus outbreak, many regions introduced stay-at-home orders. If so, ethnic fractionalization could be related to the enforcement of these restrictions, rather then voluntary social distancing as we propose.

To test this alternative explanation, we explicitly account for the introduction of the restrictive measure by regional governments. In Table B.4, we introduce both the dummy for the report of the first case of COVID-19 and the dummy for the enactment of the local stay-at-home orders. Though the introduction of stay-at-home measures led to a clear increase in social distancing, no differential effect of the stay-at-home measures exists in places with high versus low ethnic fractionalization. Unfortunately, we do not have a convincing instrument for the stay-at-home measures, so we report only the results of the OLS estimation.

Other robustness checks. We conduct several additional checks to ensure that our empirical results are robust. First, we report that when we use two-way clustering by region and day, and our results only get stronger in terms of statistical significance (Table B.5). Thus, the results reported throughout the paper are based on a more conservative specification. Second, we show that our results are robust to the inclusion of more flexible controls, e.g., baseline controls interacted with week or day fixed effects (Table B.6). The results remain nearly identical in magnitude and statistical significance. We also control for the number of cases and their interaction with ethnic fractionalization (Table B.7).[17] We find that even though the number of cases in a region increases social distancing, the differential effect of ethnic diversity is primarily driven by having a first case in the region. In addition, we control for government medical spending in a city's region; the estimates in Table B.8 show that doing so does not change our results. This addresses the possible concern that people in ethnically diverse areas may self-isolate more because of underfunded local healthcare system and higher personal cost of getting infected. Finally, we check that our results are robust to the inclusion of various measures of social capital and other-regarding preferences (Table B.9). The magnitudes and statistical significance of the ethnic fractionalization interaction remain close to our baseline results.

United States. To make sure that the results are not Russia-specific, we tested the main hypothesis that ethnic fractionalization led to a bigger reduction in mobility following the first local COVID-19 case using the United States county-level data.

[17] We can estimate this specification only with the OLS.

Table 4
Social Distancing, First Case, and Xenophobia.

	Yandex Isolation Index							
	OLS				2SLS			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Post First Case x Xenophobic Searches	0.051***	0.023**	0.051***	0.021**	0.050***	0.022*	0.051***	0.021*
	[0.011]	[0.010]	[0.011]	[0.010]	[0.012]	[0.012]	[0.012]	[0.012]
Post First Case x Ethnic Fractionalization	0.367***	0.306***	0.366***	0.310***	0.343***	0.283**	0.336***	0.274**
	[0.099]	[0.077]	[0.103]	[0.091]	[0.103]	[0.123]	[0.101]	[0.118]
Post First Case	−0.075	1.333**	−0.132**	0.904	−0.193***	0.985*	−0.226***	0.883
	[0.073]	[0.536]	[0.056]	[0.616]	[0.073]	[0.570]	[0.069]	[0.562]
Observations	17,640	17,640	17,640	17,640	17,640	17,640	17,640	17,640
R-squared	0.817	0.820	0.944	0.948	0.817	0.820	0.945	0.949
Panel B								
Post First Case x Ethnic Hate Crime	0.090***	0.032***	0.090***	0.030**	0.088***	0.032**	0.089***	0.030**
	[0.010]	[0.012]	[0.010]	[0.012]	[0.012]	[0.013]	[0.012]	[0.014]
Post First Case x Ethnic Fractionalization	0.423***	0.340***	0.425***	0.345***	0.397***	0.316***	0.390***	0.306***
	[0.109]	[0.076]	[0.108]	[0.086]	[0.101]	[0.119]	[0.098]	[0.115]
Post First Case	−0.106	1.246**	−0.164***	0.820	−0.221***	0.910	−0.255***	0.809
	[0.069]	[0.523]	[0.051]	[0.597]	[0.072]	[0.558]	[0.068]	[0.553]
Observations	17,817	17,817	17,817	17,817	17,817	17,817	17,817	17,817
R-squared	0.818	0.820	0.946	0.948	0.818	0.820	0.946	0.949
City Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes			Yes	Yes		
Day Fixed Effects			Yes	Yes			Yes	Yes
Additional conrols		Yes		Yes		Yes		Yes

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Robust standard errors in brackets are clustered by region. In columns (5)–(8), bootstrapped standard errors are reported. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020. Data captured by Yandex on xenophobic Internet searches by city was collected in 2018. Data on ethnic hate crime by city comes from NGO SOVA (2008–2015). Additional controls include interactions of the dummy for post-first-case with measures of education attainment, average wage, and population density.

Unfortunately, in the U.S., the virus spread from multiple initial sources, which makes it impossible to use an instrument similar to ours. Reassuringly, however, the OLS results, documented in Appendix D, are very similar.

7. Conclusion

This paper highlights the role of ethnic diversity in voluntary adherence to socially beneficial norms, such as self-isolation and social distancing during a pandemic. We show that people in more diverse places were more likely to restrict their mobility following the reports of the first local COVID-19 cases. While the Russian data allow us to establish a causal relation more cleanly than the data from the United States does, our results are reassuringly consistent for both countries. Theoretically, we argue that these results can be explained with a model where sick people self-isolate for altruistic reasons but do so less in more diverse societies due to out-group biases. At the same time, the decision by healthy individuals to self-isolate is motivated by private benefits, so they are more likely to self-isolate in more diverse societies, where sick people are less likely to stay at home. As long as most people consider themselves healthy, the second effect will dominate, and, on average, there will be more voluntary social distancing in more diverse societies.

Our study has important implications for government policy. We highlight not only that the propensity of different groups of people (ethnic or social groups, or healthy as opposed to sick) to engage in prosocial behavior may differ but also that there may be important strategic effects. In the context of the pandemic, decisions by healthy and sick individuals to self-isolate are strategic substitutes. This means, for example, that in a homogeneous society with high levels of tolerance, extensive testing would allow people to learn that they are sick and self-isolate, enabling the rest to go out with little fear. In a heterogeneous society with low levels of tolerance, the same policy may spur people who learn that they are contagious to go out more because they have little to lose, with the exact opposite implications for the healthy population.

There are implications for optimal strategies on reopening the economy as well. As long as most people are not sick, we expect our results to hold even after the stay-at-home orders are lifted and the extrinsic motivation to stay at home becomes weaker. Naturally, the expectation that people will voluntary observe social distancing guidelines is likely to be one of the key elements of these strategies. As long as people observe the guidelines, even in the absence of restrictive government policies, the economy can be restarted even before pharmaceutical or technological solutions to the coronavirus problem are found. These expectations, however, depend, in particular, on local ethnic diversity, and therefore reopening strategies should as well. More broadly, understanding the effects of government regulations in heterogeneous societies has practical importance beyond the pandemic, which makes it an important topic for future research.

Appendix A. Data

A.1. Social distancing indicators

For the main measure of people's movements in Russia, we use daily averages of the Yandex Isolation Index, which aggregates data on people's movements at the city level, extracted from various Yandex applications.[18] The index resembles the Google

[18] Yandex is the largest telecom company in Russia, and its main website yandex.ru is the most visited website in Russia and the twelfth-most-visited website in the world (https://www.similarweb.com/top-websites/russian-federation). The company offers many diverse products to its customers and claims to be the Russian Google, Amazon, Uber, and Spotify—all at the same time (https://www.datacenterdynamics.com/en/analysis/cloud-russia/). The company's multiple mobile apps include a web browser, a search engine, a map app, a traffic-monitoring app, an Uber-type ride-hailing platform, and a mobile payment app. The data come from https://yandex.ru/maps/covid19/isolation.

Mobility Index[19] or the Baidu's data on mobility in China (Xiao, 2020). The index is calibrated for each city by comparing the level of mobility in each particular day with the level of mobility in a typical working day before the epidemic. As a typical day, the index takes the average from Monday, March 2, through Thursday, March 5. If it is the same as during rush hour on a typical weekday, it means that the level of self-isolation is low, 0 points. If a city is as quiet as at night, this is 5 points. The higher the score, the harder it is for the virus to spread. Importantly, since the base of comparison (i.e., activity during the typical workdays) remains constant throughout the period of observation, the level of mobility is comparable over time.[20]

We use daily data for 302 cities with a population over 50,000 from February 23, 2020 through April 21, 2020. Fig. B.2 in the Appendix plots these data over time and shows that people's movements began declining even before March 29, when the first stay-at-home order was issued. In our subsequent analysis, we exclude the data on Moscow and Saint Petersburg from our sample, as these are clear outliers in many respects.

Fig. B.3 displays spatially the change in a city's isolation score before and after the first case in the city's region. Darker color indicates that the city's average isolation score went up more significantly, and that the magnitude of that change was in the higher quartile of the corresponding distribution. We present this figure alongside Fig. B.4, which spatially displays the city's quartile of ethnic fractionalization. These two maps illustrate the raw correlation between the two variables, equal to 0.4, which we will investigate more rigorously in Section 5.

A.2. Data on COVID-19 cases

We take statistics on the daily number of coronavirus cases by region from the government website that contains official information about the pandemic and policies enacted by the Russian government to fight it.[21] Fig. B.5 reports the distribution of the first case dates in our data.

Importantly for our identification strategy, though COVID-19 has reached across the country, it began spreading in Moscow (the first confirmed case was a traveler from Italy who arrived in Moscow on March 1);[22] the nation's capital accounted for more than half of all cases in Russia during the period we analyze. Fig. B.6 depicts total coronavirus cases in Russia and in Moscow from March 1 to April 30, 2020.

A.3. Other data

Migration. Our data on cross-regional migration come from the Russian Federal State Statistics Service, *RosStat*. For our empirical exercise, we distinguish between early migration (1990–1997, before the crisis of 1998) and recent migration (2015–2018). In all years, migration to Moscow, as summarized in Fig. B.7, is smaller than 10% of overall internal migration.

Xenophobia. We use two alternative measures of xenophobia in a city: online searches and the number of hate crimes. The first measure is based on the relative numbers of explicitly xenophobic Internet searches on Yandex WordStat [which is similar to Google's Search Volume Index (SVI)].[23] The data are analogous to the search-based measures of xenophobia or racism used increasingly in the literature (Stephens-Davidowitz, 2014; Ross, 2015; Chetty et al., 2019). The second measure is based on the city-level data on ethnic hate crime from the database compiled by the SOVA Center for Information and Analysis.[24] This independent, Moscow-based nonprofit organization provides information related to hate crimes; it is generally considered the most reliable source of information on this topic. The dataset covers incidents of hate crimes and violent acts of vandalism, as well as convictions under any article of the Criminal Code related to "extremism." These data have been collected consistently since 2007, with incomplete data for 2004 to 2006. In our analysis, we use 2007–2015 data. We classify all hate crimes as "ethnic" or "nonethnic" based on the type of victim reported in the database.[25]

Other Data. City-level data on population, age, education, and ethnic composition come from the Russian Censuses of 2002 and 2010. Data on the average wage and municipal budgets come from RosStat. Additional city characteristics (latitude, longitude, year the city was founded, and locations of administrative centers) come from the national encyclopedia of Russian cities and regions.[26] Table B.1 presents the summary statistics of all variables used in our analysis of the Russian case.

Appendix B. Additional evidence

See Figs. B1.1–B.8
See Tables B.1–B.10

Appendix C. Theoretical framework

C.1. Setup

Here, we provide a more detailed version of the model in Section 3, including the proofs.

Consider a simple one-period model. Society is a unit continuum of individuals G that consists of two ethnic groups G_1 (share $g_1 \in (0, \frac{1}{2}]$) and G_2 (share $g_2 = 1 - g_1$). As the game begins, each individual may be either healthy (subset H), sick (subset S), or an asymptomatic carrier (subset C). These health statuses are mutually exclusive, and the shares of healthy (h), sick (s), and carrier (c) individuals are all positive and sum to 1; furthermore, we assume that health status is independent of ethnicity. We will denote infected people as $I = C \sqcup S$ and individuals who do not exhibit symptoms as $N = H \sqcup C$. In other words,

$$G = Group_1 \sqcup Group_2 = \overbrace{Healthy \sqcup \underbrace{Carrier}_{}}^{No\ symptoms} \underbrace{\sqcup Sick}_{Infected}.$$

We introduce the group of asymptomatic carriers C both because their existence is realistic and because this allows us to obtain comparative statics with respect to the threat of asymp-

[19] https://www.google.com/covid19/mobility/

[20] For more information on the index see https://yandex.ru/company/researches/2020/podomam (in Russian).

[21] The source of data is *Rospotrebnadzor*, the government agency responsible for epidemiological surveillance. Because the website does not report historical information, we obtain those data from the Yandex coronavirus page, which uses this website as a source.

[22] Earlier, four Russians were diagnosed with coronavirus, three from the Diamond Princess cruise ship and one air passenger transiting from Iran to Azerbaijan. In addition, two Chinese citizens visiting Russia were diagnosed with COVID-19 on January 31st, but they were quickly isolated without further documented spread.

[23] There are two main differences between Google SVI and Yandex WordStat. First, the Yandex measure shows the relative numbers of searches per city, even if their absolute numbers are small. In fact, Yandex does not have a minimum number of searches before the statistics appear—even a single search is shown. Second, the Yandex measure is easily available at the city level, while Google SVI does not report city-level searches for most requests in Russia.

[24] The database can be found at https://www.sova-center.ru/en/database/

[25] More details are available in Bursztyn et al. (2019).

[26] Available at http://www.mojgorod.ru/.

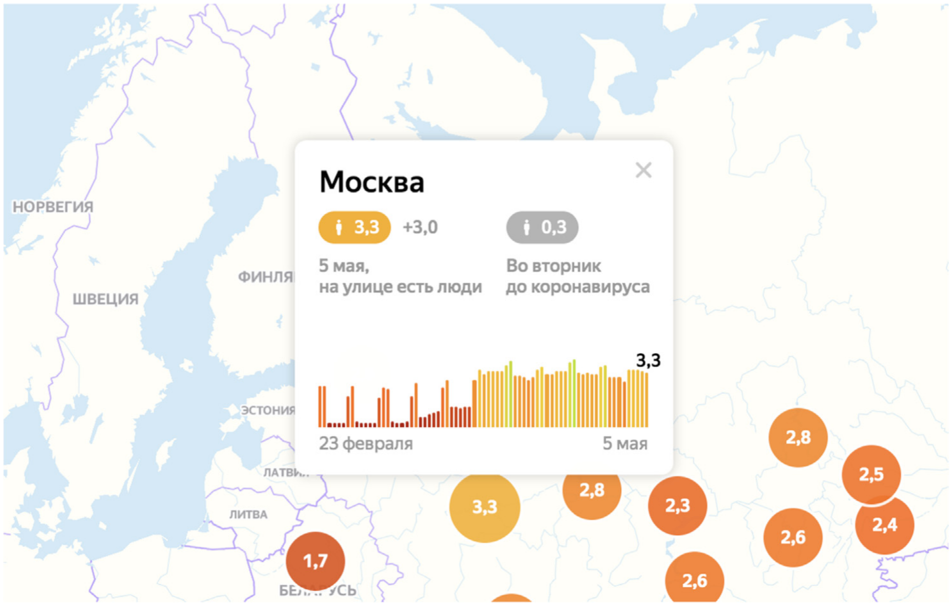

Fig. B.1. Yandex Isolation Index for Moscow, February 23 to May 5, 2020. *Source*: Yandex 2020.

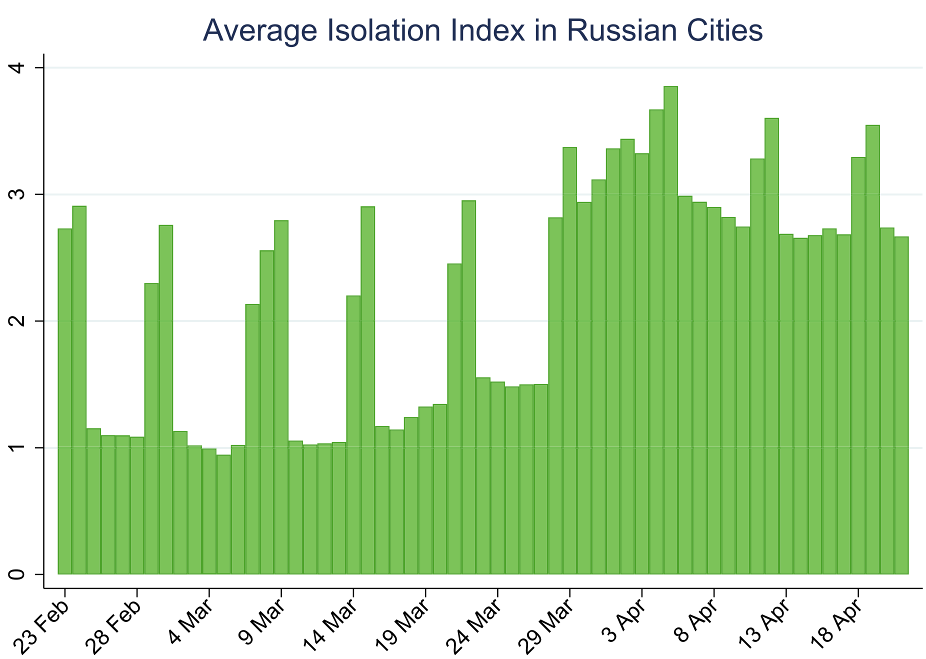

Fig. B.2. Average Isolation Index, Russian Cities, February 23 to April 21, 2020. *Notes*: Isolation index is averaged across all Russian cities with population above 50,000 people. Data come from Yandex. Monday March 9 was am official holiday, because March 8, International Women's Day, fell on a Sunday.

tomatic transmission, however, our main results will go through without it. In SubSection C.3 below, we analyze a simplified version of the model where c is assumed to equal 0.

Individuals observe whether or not they are sick, i.e., i knows if $i \in S$ or $i \in N$. However, if they do not exhibit symptoms ($i \in N$), they do not know if they are healthy ($i \in H$) or are asymptomatic carriers ($i \in C$). With this information in hand, all individuals make, simultaneously and independently, a binary decision $d_i \in \{0, 1\}$, where 1 is interpreted as self-isolation and 0 as refusal to do so (i.e., going out). Self-isolation does not produce any direct costs or benefits. Going out has a direct benefit b_i; we assume that $b_i \sim U[0, W_N]$if $i \in N$ and $b_i \sim U[0, W_S]$if $i \in S$ (and it is independent from ethnicity). It might be natural to think that $W_S \leqslant W_N$, as sick individuals may have less desire to go out, but nothing substantive changes in the model if we assume $W_S = W_N = W$. The cost of going out depends on one's health status. A healthy person may become infected, and anyone who is infected at the end of the period gets disutility $-L$ (where $L > 0$). An infected person might infect someone else, leading to a psychological cost $M > 0$ per each healthy person infected as long as this person is from the same ethnic group; the cost of infecting an out-group person is tM, where $t \in [0, 1]$captures tolerance toward individuals from the other ethnic group (i.e., lack of a negative out-group bias).

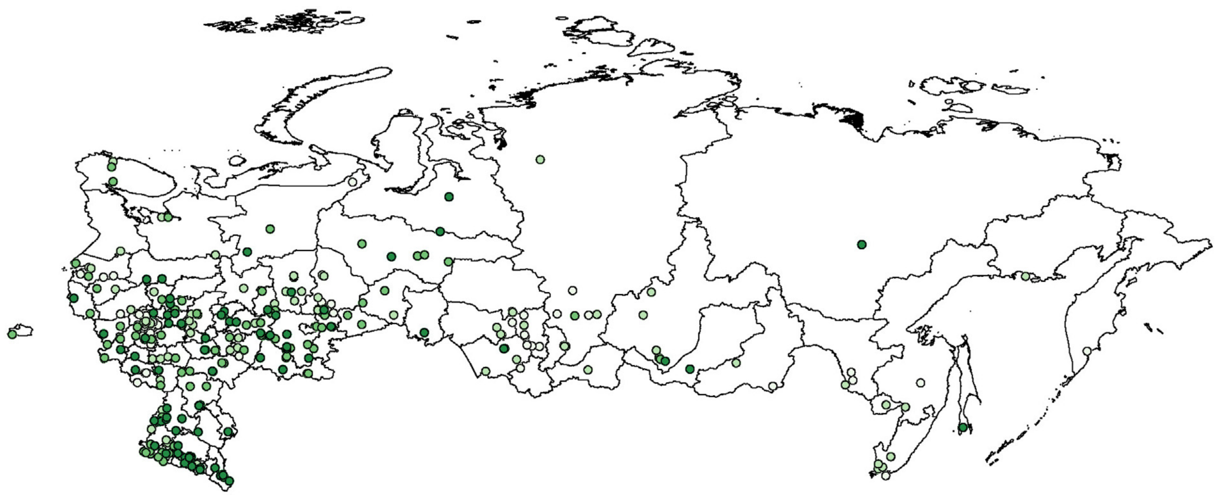

Fig. B.3. Change in Average Isolation Index of Russian Cities, Before and After First Case. *Notes*: Cities are split into four quartiles by the change in their average isolation index before and after the first case in the cities' regions. Darker color indicates greater change in the isolation index. The period is from February 23 to April 21, 2020. The sample includes 302 Russian cities with a population of at least 50,000. Black lines within Russia demarcate regional borders. Chukotka Autonomous Okrug is omitted for illustration purposes.

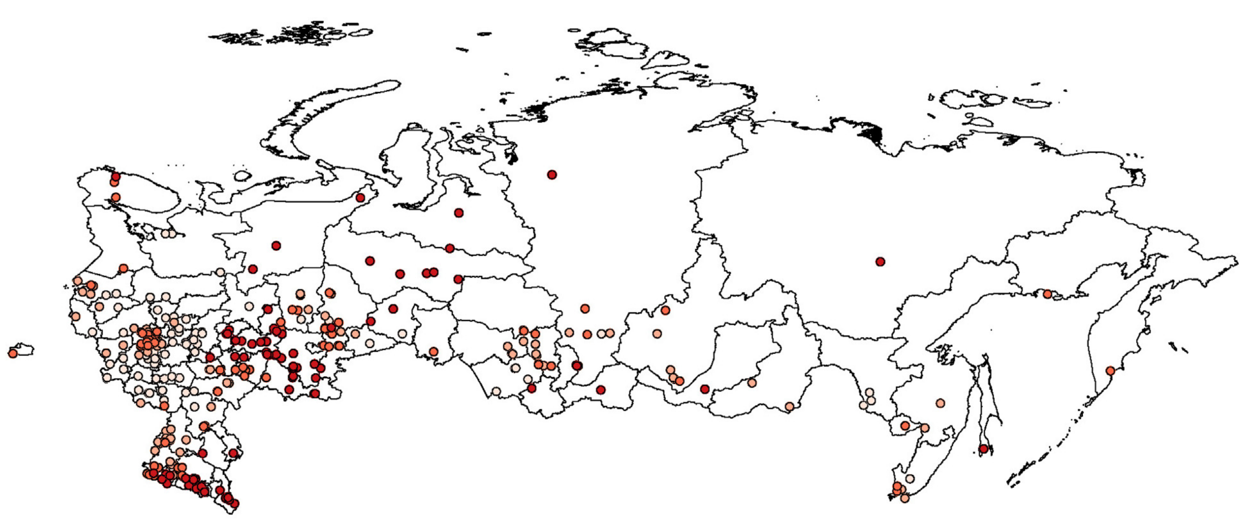

Fig. B.4. Ethnic Fractionalization in Russian Cities. *Notes*: Cities are split into four quartiles by ethnic fractionalization per 2010 Census. Darker color indicates higher fractionalization. The sample includes 302 Russian cities with a population of at least 50,000. Black lines within Russia demarcate regional borders. Chukotka Autonomous Okrug is omitted for illustration purposes.

Consider the following simplified model of interactions during a pandemic. Suppose that all individuals are matched in pairs, and let $m(i)$ denote the match of individual i. Assume that if everyone goes out, then each i would come in close contact with exactly one other person, their match $m(i)$. If one or both of two matched individuals decide to stay home, there would be no transmission of the infection between them. The same is true if they both go out, if both are healthy or both are infected (regardless if they are carriers or are sick). If one is healthy and the other is infected, the healthy one becomes infected with probability q if the infected person is sick and r if the infected person is a carrier.[27] Naturally, $r > 0$ reflects the possibility of asymptomatic transmission.

When deciding whether to self-isolate or not, individuals do not know who they are matched with, but they know the distribution of types. Thus, individuals who show no symptoms ($i \in N$) choose d_i to maximize their expected utility:

$$U_N = -\frac{c}{h+c}L + b_i \mathbf{1}_{d_i=0} - \frac{h}{h+c}\left(q\mathbf{1}_{m(i)\in S} + r\mathbf{1}_{m(i)\in C}\right)L\mathbf{1}_{d_i=d_{m(i)}=0} - \frac{c}{h+c}r\mathbf{1}_{m(i)\in H}\left(M\mathbf{1}_{G(i)=G(m(i))} + tM\mathbf{1}_{G(i)\neq G(m(i))}\right)\mathbf{1}_{d_i=d_{m_i}=0}, \quad \text{(C1)}$$

while sick individuals ($i \in S$) maximize

$$U_S = -L + b_i\mathbf{1}_{d_i=0} - q\mathbf{1}_{m(i)\in H}\left(M\mathbf{1}_{G(i)=G(m(i))} + tM\mathbf{1}_{G(i)\neq G(m(i))}\right)\mathbf{1}_{d_i=d_{m(i)}=0}. \quad \text{(C2)}$$

We are interested in the Perfect Bayesian equilibria of this game. To focus on the interesting case, we maintain the following assumption:

[27] The probability of getting infected is thus proportional to the mass of infected individuals who go out, weighted by their contagiousness. In practice, this relationship may be more complex. For example, it may be concave because of the possibility of getting infected by multiple individuals, or it might be convex, for example, because close interactions are easier to avoid when few sick people are out. We adopt the simple proportionality assumption for simplicity.

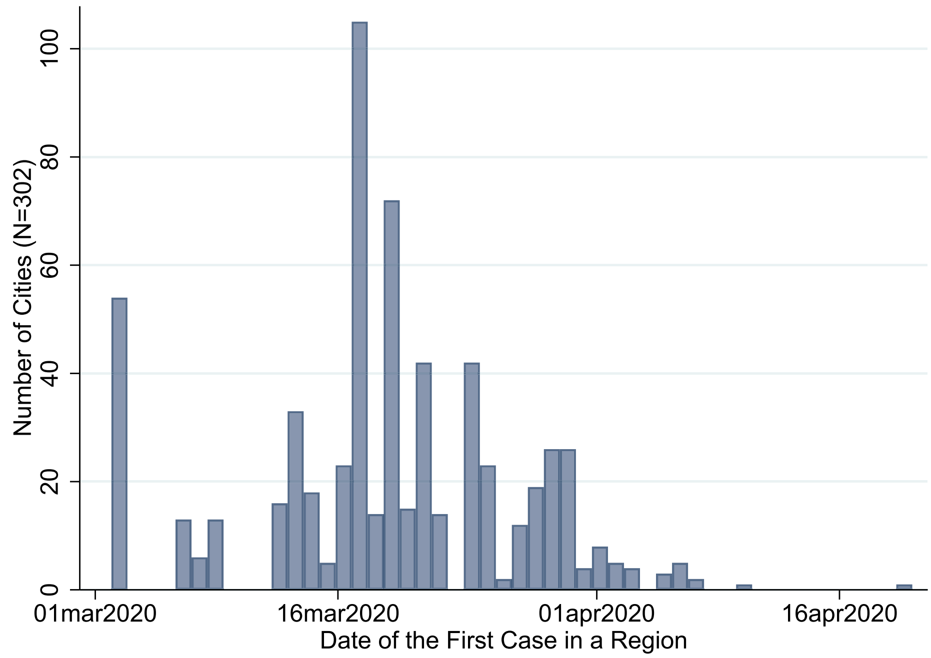

Fig. B.5. Distribution of Dates of the First Case of Coronavirus by Region (Excluding Moscow). *Notes*: The sample is all Russian cities with population above 50,000 people. Data come from RosPotrebNadzor.

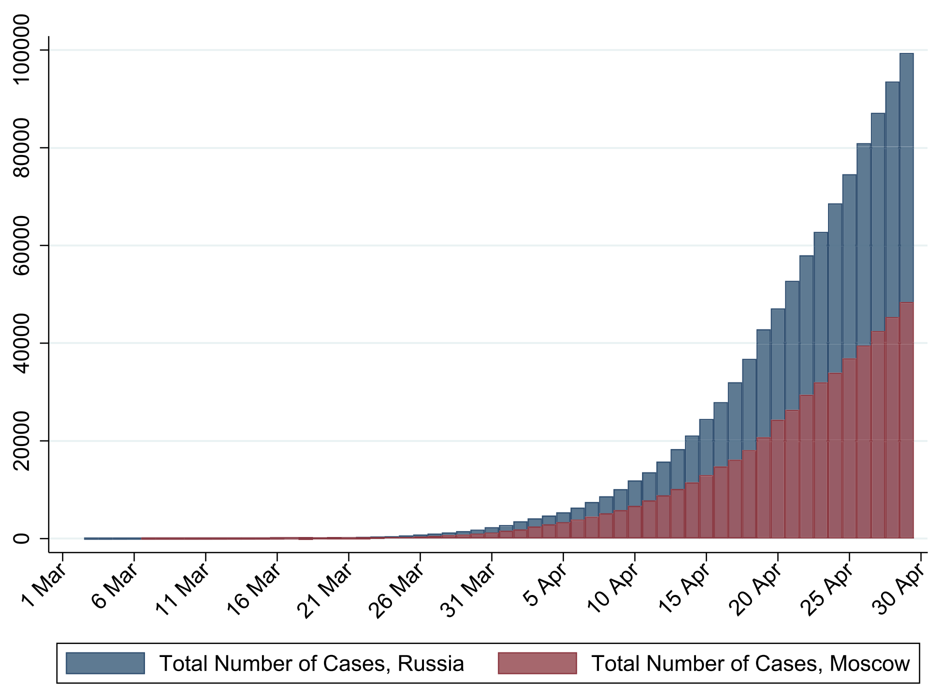

Fig. B.6. COVID-19 Cases, March 1 to April 30, 2020, Russia and Moscow. *Source*: RosPotrebNadzor.

Assumption C1. $W_N < qL$ and $W_S > qM$.

This first part of the condition is satisfied if the disutility of getting infected L is high enough. Specifically, it states that if a healthy person were certain to encounter a sick individual (and thus get infected with probability q), this person would prefer to stay home. The second condition suggests that altruism M is not too high. This condition means that at least some sick individuals (those with b_i sufficiently high) would go out even if they were certain to encounter a healthy individual. If this condition were to fail, altruism would keep all sick individuals at home, at least when most people are healthy. This upper boundary on M also happens to be sufficient (though not necessary) to guarantee existence and uniqueness of an equilibrium.

C.2. Analysis

A Perfect Bayesian equilibrium of this game is characterized by four cutoffs, $\beta_{N1}, \beta_{N2} \in [0, W_N]$and $\beta_{S1}, \beta_{S2} \in [0, W_S]$, such that individual i with health status $j \in \{N, S\}$ from ethnic group

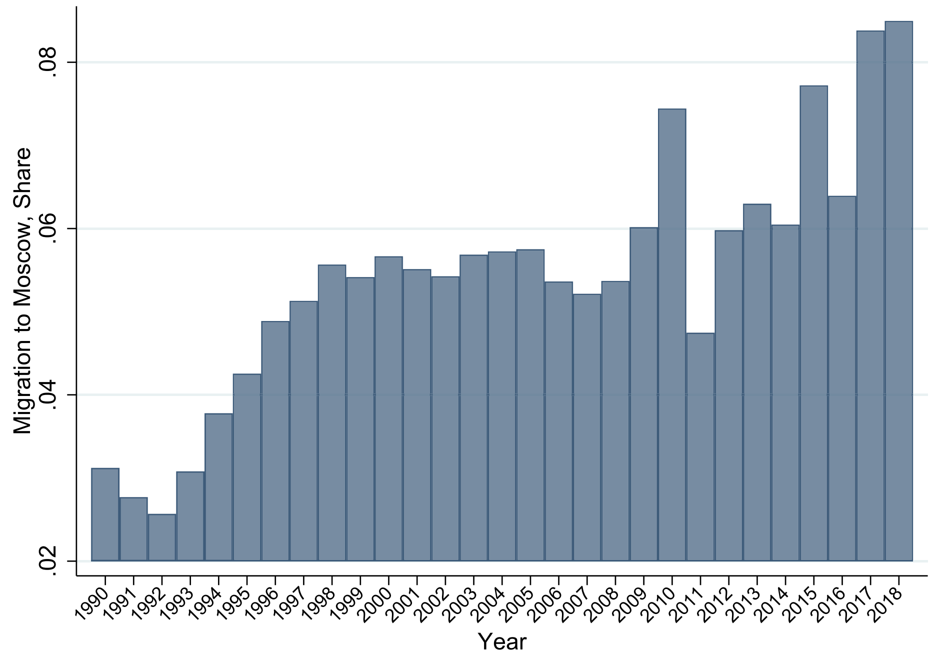

Fig. B.7. Migration to Moscow as a Share of Inter-regional Migration, 1990–2018. *Source*: *RosStat*.

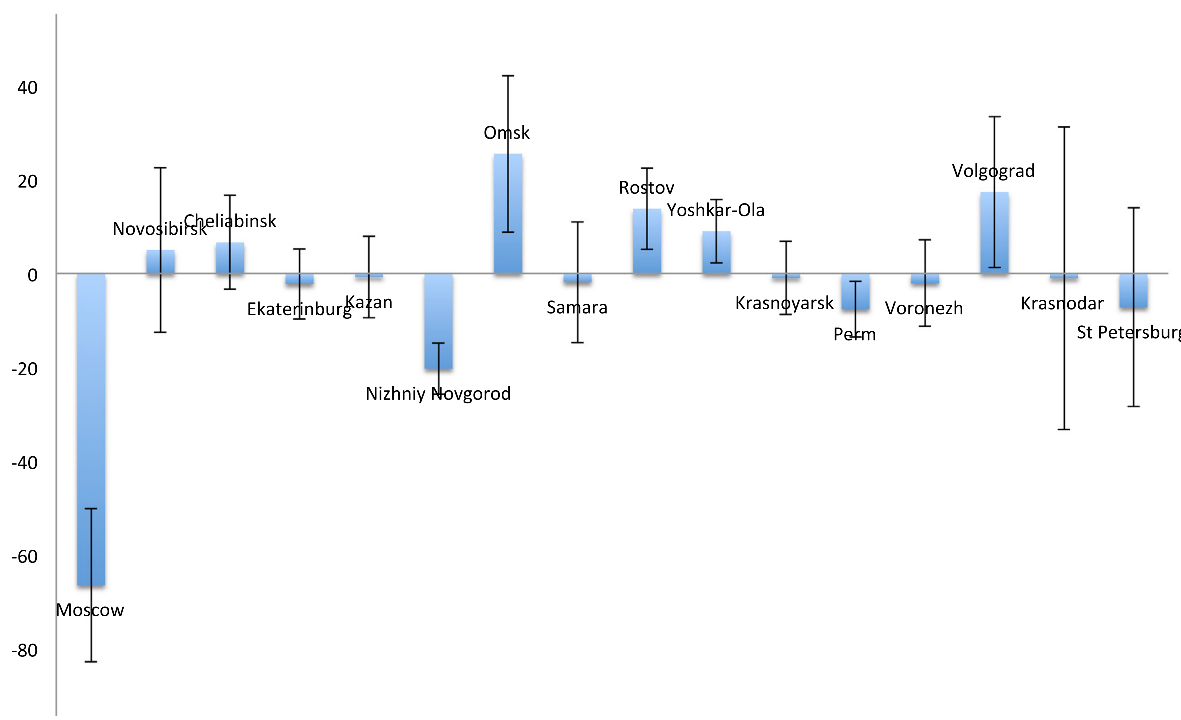

Fig. B.8. Distribution of the First-Stage Coefficients for Various Cities.*Notes*: The reported coefficients are for the share of outmigration to a given region (the key independent variable in Tables 2 and B.3).

$G_k, k \in \{1, 2\}$, self-isolates if $b_i < \beta_{jk}$ and goes out if $b_i > \beta_{jk}$. The following proposition characterizes the equilibrium:

Proposition C1. *If $W_N > q\frac{h}{h+c}sL$, there is a unique interior equilibrium, in which $0 < \beta_{N1} < \beta_{N2} < W_N$ and $0 < \beta_{S1} < \beta_{S2} < W_S$ (provided that $g_1 < \frac{1}{2}$ and $t < 1$). Otherwise, in the unique equilibrium, $\beta_{N1} = \beta_{N2} = W_N, \beta_{S1} = \beta_{S2} = 0$, so all people without symptoms self-isolate and all sick people go out.*

The coefficient $q\frac{h}{c+h}s$ in the first condition is the probability that a person without symptoms will get infected by a sick person if all sick people go out. If this probability is sufficiently low, then at least some people without symptoms will go out (the first person to do so will not be afraid of getting infected by another such person, so the possibility of asymptomatic transmission does not enter this condition). For example, this condition is guaranteed to hold if $s = 0$, i.e., in the beginning of the pandemic. In equilibrium, people

Table B.1
Summary Statistics—Russia.

VARIABLES	Obs.	Mean	S.D.	Min	Max
Yandex isolation index	17,817	2.26	0.98	0.10	4.50
Number of COVID-19 cases	17,817	113.6	546.4	0	5959
Ethnic fractionalization, 2010	17,817	0.22	0.18	0.01	0.85
% with higher education, 2010	17,817	0.27	0.08	0.13	0.46
Average wage, 2011	17,817	9.97	0.33	9.18	11.09
Population density	17,817	17.06	13.27	0.10	102.60
Share of migration to Moscow, 2015–2018	17,817	0.09	0.09	0.01	0.30
Predicted share of migration to Moscow, shift-share IV	17,817	0.01	0.01	0.00	0.03
Xenophobic searches	17,640	0.81	1.14	−2.27	4.13
Log of total number of ethnic crimes, 2008–2015	17,817	0.66	1.01	0.00	3.99
Social web searches, principal component	17,345	−0.07	1.57	−3.32	10.22
NGO per capita	17,817	0.70	0.50	0.00	3.05
Trust, 2012	17,758	0.70	0.10	0.37	0.85
Social capital, principal component	17,345	0.23	1.15	−2.15	5.82

Table B.2
Summary Statistics—United States.

VARIABLES	Obs.	Mean	S.D.	Min	Max
% staying home	360,870	26.42	8.51	1.38	80.77
Number of COVID-19 cases in a county	360,870	30.22	432.10	0	32124
Number of COVID-19 cases in a state	360,870	2,423	13,110	0	263,460
Number of COVID-19 deaths in a county	360,870	1.08	18.43	0	1813
Number of COVID-19 deaths in a state	360,870	94.42	673.40	0	15,740
Ethnic fractionalization, 2010	360,870	0.29	0.19	0.017	0.762
% of adults with a BA degree	360,870	20.80	9.14	3	80.2
Median HH income ('000s)	360,755	47.98	12.59	18.97	125.70
Population density	360,870	0.268	1.788	3.9E-05	71.62
Average Turnout, 2012 & 2016	360,870	59.2	9.553	13.13	92.46
Charity Donations, Share of AGI, middle-class itemizers	358,915	4.454	1.836	0	18.7
PSU Social Capital Index	360,870	0.002	1.260	−3.183	21.81
SCP Social Capital Index	343,850	0.005	1.003	−4.315	2.971

Table B.3
Timing of First Case and Internal Migration to Other Large Cities, 2015–2018.

	Date of the First Covid-19 case in a Region			
	OLS		IV	
VARIABLES	(1)	(2)	(3)	(4)
Migration to Other Large Cities in 2015–2018	0.631	−3.942	3.684	−0.259
	[6.923]	[4.147]	[7.658]	[4.897]
Average Wage		−7.323**		−7.201**
		[3.434]		[3.414]
Education		2.392		5.211
		[10.420]		[10.434]
Population Density		−0.251***		−0.237***
		[0.050]		[0.051]
Observations	302	302	302	302
R-squared	0.000	0.211	−0.009	0.199
Kleibergen- Paap			198.5	183.1

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Robust standard errors in brackets are clustered by region. The sample includes 302 Russian cities with a population of at least 50,000. Migration to Other Large Cities is computed as the aggregate migration to regions with cities with a population of at least 1 million. The list of regions consists of Novosibirskaya oblast, Chelyabinskaya oblast, Sverdlovskaya oblast, Tatarstan Republic, Nizhegorodskaya oblast, Samarskaya oblast, Rostovksaya oblast, Bashkortostan Republic, Krasnoyarskyi krai, Permskyi krai, Voronezhskaya oblast, Volgogradskaya oblast, and Krasnodarsky krai. In columns (3) and (4), migration to these regions is predicted with a shift-share instrument, using pre-1998 migration to these regions combined with 2015–2018 aggregate outflow of internal migration from a source region.

from the ethnic minority are less likely to self-isolate, because the person they might infect is likely to be from the majority group, whereas the probability of getting infected is the same for healthy individuals from both ethnic groups.

We now turn to the comparative statics results.

Proposition C2. *Suppose that $W_N > q\frac{h}{h+c}sL$, so the equilibrium is interior. Then an increase in the size of the minority group g_1, a decrease in altruism M, or a decrease in tolerance t all decrease self-isolation by sick individuals. The effect on overall self-isolation is ambiguous: it increases as a result of either of these changes if $\frac{rc}{qs} < \frac{qhL-W_N}{W_S+q\frac{h}{h+c}sL}$, and it decreases if the converse is true.*

In the light of Assumption C1, the right-hand side of the last condition is positive for h close to 1, i.e., in the beginning of the pandemic. This means that the comparative statics critically depends on the likelihood of asymptomatic transmission (and the share of asymptomatic people) relative to the

Table B.4
Social Distancing, First Case, and Stay-at-Home Orders.

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	0.281**	0.235**	0.284***	0.240**
	[0.109]	[0.105]	[0.093]	[0.117]
Post First Case	−0.005	1.214**	−0.070	0.803
	[0.070]	[0.472]	[0.055]	[0.562]
Stay at Home Measures x Ethnic Fractionalization	0.033	0.019	0.076	0.064
	[0.191]	[0.190]	[0.145]	[0.140]
Stay at Home Measures	0.362**	0.353**	0.197***	0.187***
	[0.167]	[0.158]	[0.074]	[0.060]
Post First Case x Education		1.822***		1.786***
		[0.256]		[0.260]
Post First Case x Average Wage		−0.174***		−0.139**
		[0.050]		[0.060]
Post First Case x Population Density		0.003**		0.003**
		[0.001]		[0.001]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,817	17,817	17,817	17,817
R-squared	0.819	0.823	0.945	0.949

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Robust standard errors in brackets are clustered by region. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020.

Table B.5
Social Distancing, First Case, and Ethnic Fractionalization (2SLS). Alternative Clustering.

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post Predicted First Case x Ethnic Fractionalization	0.352***	0.293***	0.345***	0.285***
	[0.032]	[0.037]	[0.024]	[0.023]
Post Predicted First Case	−0.154***	0.893***	−0.186***	0.793***
	[0.013]	[0.203]	[0.010]	[0.126]
Post Predicted First Case x Education		1.798***		1.813***
		[0.086]		[0.065]
Post Predicted First Case x Average Wage		−0.156***		−0.151***
		[0.021]		[0.013]
Post Predicted First Case x Population Density		0.003***		0.003***
		[0.001]		[0.000]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,817	17,817	17,817	17,817
R-squared	0.816	0.820	0.944	0.949

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Bootstrapped robust standard errors in brackets are clustered by region and day. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020. Predicted First Case is computed using the data on inter-regional migration, as summarized above

Table B.6
Social Distancing, First Case, and Ethnic Fractionalization (2SLS). Flexible Controls.

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post Predicted First Case x Ethnic Fractionalization	0.302***	0.295***	0.298***	0.296***
	[0.114]	[0.113]	[0.112]	[0.114]
Post Predicted First Case	−0.063	−0.080	−0.086*	−0.081
	[0.052]	[0.050]	[0.051]	[0.053]
Baseline Controls Interacted with Week Fixed Effects	Yes		Yes	
Baseline Controls Interacted with Day Fixed Effects		Yes		Yes
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,817	17,817	17,817	17,817
R-squared	0.816	0.820	0.944	0.949

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Bootstrapped robust standard errors in brackets are clustered by region and day. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020. Predicted First Case is computed using the data on inter-regional migration, as summarized above. Baseline controls include population density, average wage, and average share of those with higher education.

Table B.7
Social Distancing, First Case, # of Cases, and Ethnic Fractionalization (OLS).

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	0.436***	0.362**	0.381***	0.309*
	[0.132]	[0.154]	[0.139]	[0.167]
Log (# or Covid-19 cases) x Ethnic Fractionalization	−0.023	−0.019	−0.005	−0.001
	[0.041]	[0.042]	[0.038]	[0.038]
Post First Case	−0.075	1.590***	−0.129***	1.217**
	[0.057]	[0.524]	[0.036]	[0.588]
Log (# or Covid-19 cases)	0.107***	0.098***	0.116***	0.106***
	[0.020]	[0.018]	[0.017]	[0.016]
Post First Case x Education		1.879***		1.818***
		[0.258]		[0.258]
Post First Case x Average Wage		−0.215***		−0.181***
		[0.057]		[0.064]
Post First Case x Population Density		−0.000		−0.000
		[0.002]		[0.002]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,817	12,803	17,817	12,803
R-squared	0.816	0.819	0.944	0.950

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Bootstrapped robust standard errors in brackets are clustered by region and day. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020.

Table B.8
Social Distancing, First Case, and Government Medical Spending (2SLS).

	Yandex Isolation Index			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	0.363***	0.285***	0.358***	0.276***
	[0.111]	[0.101]	[0.106]	[0.096]
Post First Case	−2.756	−0.929	−3.093	−1.239
	[1.889]	[1.420]	[1.986]	[1.373]
Post First Case x Log Government Medical Spending (p.c.)	0.274	0.314	0.306	0.348*
	[0.195]	[0.191]	[0.205]	[0.186]
Post First Case x Education		1.868***		1.890***
		[0.259]		[0.256]
Post First Case x Average Wage		−0.272***		−0.279***
		[0.075]		[0.075]
Post First Case x Population Density		0.002		0.002*
		[0.002]		[0.001]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Observations	17,699	17,699	17,699	17,699
R-squared	0.817	0.820	0.945	0.948

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Robust standard errors in brackets are clustered by region. Isolation index, the aggregate measure of staying at home, is based on mobile app data. Logarithm of government medical spending per capita is measured at the regional level. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020.

likelihood of transmission from sick people (and their share). If this ratio is small, then higher fractionalization implies less self-isolation by sick individuals, but more self-isolation overall, because healthy individuals are concerned about getting infected by sick people who self-isolate less. If, however, asymptomatic transmission is a major issue, then higher fractionalization also means that people without symptoms are less concerned about infecting healthy ones, and thus overall self-isolation may decrease. As h becomes small (e.g., later in the pandemic), the comparative statics becomes driven solely by sick individuals, and fractionalization will imply less self-isolation. The effect of a decrease in altruism or tolerance is similar.

Proposition C2 implies, in particular, that we should expect fractionalization to have a positive effect on self-isolation in the beginning of the pandemic (h close to 1) and in cases where asymptomatic transmission is believed to be impossible or considerably less likely than transmission from sick individuals ($\frac{r}{q}$ close to 0). Of course, in the extreme, if $h = 1$ (i.e., before the pandemic), there is no self-isolation, and this does not depend on fractionalization or tolerance.

C.3. No asymptomatic carriers

The analysis becomes particularly simple yet insightful if there are no asymptomatic carriers. In this case, $C = \varnothing, c = 0$, and $s = 1 - h$. A Perfect Bayesian equilibrium of this game is characterized by four cutoffs, $\beta_{N1}, \beta_{N2} \in [0, W_N]$and $\beta_{S1}, \beta_{S2} \in [0, W_S]$, such that individual i with health status $j \in \{N, S\}$ from ethnic group $G_k, k \in \{1, 2\}$, self-isolates if $b_i < \beta_{jk}$ and goes out if $b_i > \beta_{jk}$. Here, $N = H$, so the cutoffs β_{N1} and β_{N2} refer to actions of healthy individuals. The following proposition characterizes the equilibrium, similarly to Proposition C1:

Table B.9
Social Distancing, First Case, Ethnic Fractionalization, and Social Capital (2SLS).

	Yandex Isolation Index							
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post Predicted First Case x Ethnic Fractionalization	0.268**	0.258**	0.334***	0.286**	0.353***	0.297***	0.286**	0.281**
	[0.134]	[0.116]	[0.091]	[0.118]	[0.125]	[0.109]	[0.133]	[0.111]
Post Predicted First Case	−0.175***	0.683	−0.295***	0.888	0.093	0.751	−0.182***	0.768
	[0.050]	[0.825]	[0.078]	[0.565]	[0.136]	[0.561]	[0.066]	[0.622]
Post Predicted First Case x Social Web Searches	−0.045***	−0.017**						
	[0.011]	[0.008]						
Post Predicted First Case x NGO per capita			0.159***	0.043				
			[0.037]	[0.033]				
Post Predicted First Case x Trust					−0.406**	−0.322*		
					[0.200]	[0.164]		
Post Predicted First Case x Social Capital							0.084***	0.022*
							[0.016]	[0.013]
City Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Post Predicted First Case Interacted w/ Baseline Controls		Yes		Yes		Yes		Yes
Observations	17,345	17,345	17,817	17,817	17,758	17,758	17,345	17,345
R-squared	0.945	0.949	0.946	0.949	0.945	0.949	0.944	0.949

Notes: Bootstrapped robust standard errors are in brackets. Standard errors are clustered by region. *** p < 0.01, ** p < 0.05, * p < 0.1. Isolation index, the aggregate measure of staying at home, is based on mobile app data. The sample includes 302 Russian cities with a population of at least 50,000. The period is February 23 to April 21, 2020. Predicted First Case is computed using the data on inter-regional migration, as summarized above. Regional level of trust is computed based on the large representative survey conducted by FOM (*Fond Obshchestvennogo Mneniya*) of more than 34,000 people in 2014. Trust is measured as a response to the question "Generally speaking, do you believe that most people can be trusted or you can't be too careful in dealing with people?" Social web searches are computed at the city level for Yandex web searches conducted in 2019–2020; we use them as a proxy for social capital following Guriev and Melnikov (2016). Social capital measure is the first principal component of the previous three measures.

Table B.10
Social Distancing, Ethnic Fractionalization, and Social Capital. U.S. Data.

	% Staying Home			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	4.683***	5.699***	2.456*	5.392***
	[1.346]	[1.360]	[1.340]	[1.547]
Post First Case x Turnout	0.080***			
	[0.022]			
Post First Case x Charity Donations		−0.688***		
		[0.162]		
Post First Case x PSU Social Capital Index			−0.626***	
			[0.148]	
Post First Case x SCP Social Capital Index				0.529*
				[0.277]
County Fixed Effects	Yes	Yes	Yes	Yes
Day Fixed Effects	Yes	Yes	Yes	Yes
Days	115	115	115	115
Counties	3,138	3,121	3,138	2,990
Observations	360,870	358,915	360,870	343,850
R-squared	0.788	0.796	0.788	0.808

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Standard errors in brackets are clustered at the state level. Percentage of people staying home is calculated based on the number of mobile devices never leaving house divided by the total number of mobile devices observed in the county that day. The time period is 01/01/2020–24/04/2020. Post first case indicator is equal to one after a county's state already had its first COVID-19 case, and zero otherwise. Turnout is the average presidential election voting rate in 2012 and 2016; Charity Donations are calculated as charitable contributions as share of AGI, middle-class itemizers; PSU Social Capital Index is the Penn State social capital index; SCP Social Capital Index is the Social Capital Project social capital index.

Proposition C3. If $W_N > qsL$, there is a unique interior equilibrium, in which $0 < \beta_{N1} = \beta_{N2} < W_N$ and $0 < \beta_{S1} < \beta_{S2} < W_S$ (provided that $g_1 < \frac{1}{2}$ and $t < 1$). Otherwise, in the unique equilibrium, $\beta_{N1} = \beta_{N2} = W_N, \beta_{S1} = \beta_{S2} = 0$, so all people without symptoms self-isolate and all sick people go out.

Notice that unlike Proposition C1, in Proposition C3 healthy people of both ethnicities are equally likely to self-isolate. The reason is that in the absence of asymptomatic carriers, healthy people know that they are healthy, and are therefore not concerned about infecting anyone else. Their chances of meeting a sick person and getting infected are equal, and they therefore solve exactly the same problem, which results in $\beta_{N1} = \beta_{N2}$. This simplifies the formulas considerably, and we can provide the following closed-form solution for the interior equilibrium:

$$\begin{aligned}\beta_{Nk} = &\quad W_N\left(1 - W_S\frac{W_N - qsL}{W_N W_S - q^2 hsLM(1-2g_1g_2(1-t))}\right);\\ \beta_{Sk} = &\quad W_S\frac{qhM(1-(1-g_k)(1-t))(W_N - qsL)}{W_N W_S - q^2 hsLM(1-2g_1g_2(1-t))}.\end{aligned}$$

It is straightforward to see that if $g_1 < \frac{1}{2}$, then sick individuals from group G_1 are less likely to self-isolate than those from group G_2; intuitively, members of the smaller group are less concerned about infecting a person of the same ethnicity, because they are fewer in numbers.

The comparative statics result is now stated as follows.

Proposition C4. Suppose that $W_N > qsL$, so the equilibrium is interior. Then an increase in the size of the minority group g_1, a decrease in altruism M, or a decrease in tolerance t all decrease self-isolation by sick individuals. The effect on overall self-isolation increases as a result of either of these changes if $W_N < qhL$, and it decreases if the converse is true.

In the light of Assumption C1, the last condition holds for for h close to 1, i.e., in the beginning of the pandemic. This supports the intuition of the paper: in the beginning of the pandemic, if asymptomatic transmission is unlikely (and in the particular case we are analyzing here, impossible), then an increase in ethnic fractionalization, or a decrease in altruism or tolerance all increase overall self-isolation. Thus, in this simplified model, we get similar empirical predictions, in particular, that we should expect fractionalization to have a positive effect on self-isolation in the beginning of the pandemic.

C.4. Proofs

Proof of Proposition C1. Consider the four cutoffs, $\beta_{N1}, \beta_{N2} \in [0, W_N]$and $\beta_{S1}, \beta_{S2} \in [0, W_S]$.

First, let us show that $\beta_{S1}, \beta_{S2} < W_S$ in any equilibrium. Indeed, by Assumption C1, the utility of a sick person with $b_i = W_S$ satisfies

$$U_S(d_i = 0) - U_S(d_i = 1) \geqslant W_S - qM > 0,$$

so this person strictly prefers to go out, and thus so do those with b_i close to W_S.

Second, let us show that $\beta_{N1}, \beta_{N2} > 0$. The utility of a person who shows no symptoms with $b_i = 0$ satisfies

$$U_N(d_i = 0) - U_N(d_i = 1) < 0,$$

because there is a positive probability that $m(i) \in S$ and $d_{m_i=0}$, as we just showed.

Third, let us show that $\beta_{N1} = W_N$ if and only if $\beta_{N2} = W_N$. Suppose, to obtain a contradiction, that $\beta_{N1} = W_N$, but $\beta_{N2} < W_N$. Consider two individuals, $i \in G_1 \cap N$ and $j \in G_2 \cap N$ such that $b_i = b_j = W_N$. For these individuals, the probability of getting infected if they go out are equal. The probabilities of infecting someone are equal as well, but since $\beta_{N1} = W_N$, a healthy individual who goes out belongs to G_2 with probability 1. Since $t \leqslant 1$, the payoff of individual i from going out is at least as high as that of individual j, i.e., it is positive. This contradicts the hypothesis that $\beta_{N1} = W_N$. The opposite case, where $\beta_{N1} < W_N$, but $\beta_{N2} = W_N$, is considered similarly.

Fourth, let us prove that $\beta_{S1} = 0$ if and only if $\beta_{S2} = 0$, and in that case $\beta_{N1} = \beta_{N2} = W_N$. Indeed, suppose $\beta_{S1} = 0$ (the argument in the opposite direction is similar). Then individual $i \in G_1 \cap S$ with $b_i = 0$ weakly prefers to go out, but this is only possible if he zero chance of infecting anyone. The latter is only possible if healthy people all stay home, i.e., if $\beta_{N1} = \beta_{N2} = W_N$. In that case, however, all individuals $j \in G_2 \cap S$ weakly prefer to go out, so $\beta_{S2} = 0$.

Fifth, let us prove that $\beta_{N1} = \beta_{N2} = W_N$ is part of equilibrium if and only if $W_N \leqslant q\frac{h}{h+c}sL$, and in that case $\beta_{S1} = \beta_{S2} = 0$. The latter part is trivial: if almost all individuals without healthy stay at home, then almost all sick individuals strictly prefer to go out, because infecting anyone is impossible. These strategies form an equilibrium if individuals with no symptoms from both groups indeed prefer to stay home. Consider an individual $i \in N$ from either group. For such a person,

$$U_N(d_i = 1) - U_N(d_i = 0) = \frac{h}{h+c}qsL - b_i,$$

because all sick individuals go out, and (almost) none of the healthy ones does. If $W_N \leqslant q\frac{h}{h+c}sL$, then all individuals with $b_i < W_N$strictly prefer to stay at home; otherwise, individual with $b_i = W_N$ prefers to go out. This proves that $\beta_{N1} = \beta_{N2} = W_N, \beta_{S1} = \beta_{S2} = 0$ form an equilibrium if and only if $W_N \leqslant q\frac{h}{h+c}sL$. Furthermore, from the previous steps it follows that this is the only non-interior equilibrium possible.

Sixth, consider interior equilibria. In such an equilibrium, cutoff individuals are all indifferent between the two actions, $d_i = 0$ and $d_j = 1$. This gives rise to the following system of equations (we denoted the shares of each group that go out by $\gamma_{N1} \equiv \frac{W_N - \beta_{N1}}{W_N}, \gamma_{N2} \equiv \frac{W_N - \beta_{N2}}{W_N}, \gamma_{S1} \equiv \frac{W_S - \beta_{S1}}{W_S}, \gamma_{S2} \equiv \frac{W_S - \beta_{S2}}{W_S}$, and further $\gamma_N \equiv g_1\gamma_{N1} + g_2\gamma_{N2}$ and $\gamma_S \equiv g_1\gamma_{S1} + g_2\gamma_{S2}$):

$$\begin{aligned}
\beta_{N1} &= \tfrac{h}{h+c}(s\gamma_S q + c\gamma_N r)L + \tfrac{c}{h+c}h(g_1\gamma_{N1} + g_2 t\gamma_{N2})rM \\
\beta_{N2} &= \tfrac{h}{h+c}(s\gamma_S q + c\gamma_N r)L + \tfrac{c}{h+c}h(g_1 t\gamma_{N1} + g_2\gamma_{N2})rM \\
\beta_{S1} &= h(g_1\gamma_{N1} + g_2 t\gamma_{N2})qM \\
\beta_{S2} &= h(g_1 t\gamma_{N1} + g_2\gamma_{N2})qM
\end{aligned}$$

Denoting $g_1 = g$ and $g_2 = 1 - g$, we can write in matrix form,

$$A\begin{pmatrix}\beta_{N1}\\ \beta_{N2}\\ \beta_{S1}\\ \beta_{S2}\end{pmatrix} = z,$$

where

$$A = \begin{pmatrix}
1 + \frac{hg}{c+h}c\frac{rL}{W_N} + \frac{cg}{c+h}h\frac{rM}{W_N} & \frac{h(1-g)}{c+h}c\frac{rL}{W_N} + \frac{c(1-g)}{c+h}h\frac{rtM}{W_N} & \frac{hg}{c+h}s\frac{qL}{W_S} & \frac{h(1-g)}{c+h}s\frac{qL}{W_S} \\
\frac{hg}{c+h}c\frac{rL}{W_N} + \frac{cg}{c+h}h\frac{rtM}{W_N} & 1 + \frac{h(1-g)}{c+h}c\frac{rL}{W_N} + \frac{c(1-g)}{c+h}h\frac{rM}{W_N} & \frac{hg}{c+h}s\frac{qL}{W_S} & \frac{h(1-g)}{c+h}s\frac{qL}{W_S} \\
hg\frac{qM}{W_N} & h(1-g)\frac{qtM}{W_N} & 1 & 0 \\
hg\frac{qtM}{W_N} & h(1-g)\frac{qM}{W_N} & 0 & 1
\end{pmatrix}$$

and

$$z = \begin{pmatrix}
\frac{h}{c+h}(sqL + crL) + \frac{c}{c+h}h(grM + (1-g)rtM) \\
\frac{h}{c+h}(sqL + crL) + \frac{c}{c+h}h(grtM + (1-g)rM) \\
hgqM + h(1-g)qtM \\
hgqtM + h(1-g)qM
\end{pmatrix}.$$

Let us compute the determinant of matrix A. After simplifications, we find

$$\det A = \frac{y_1 g(1-g) + y_2}{(c+h)^2 W_N^2 W_S},$$

where

$$\begin{aligned}
y_1 &= h^2(1-t)M\Big(2q^2s(c+h)LW_N + c^2r^2(2L + (1+t)M)W_S \\
&\quad - chq^2rs(t+1)LM\Big) \\
y_2 &= W_N(c+h)\Big((cW_N + hW_N + chrL + chrM)W_S - LMh^2q^2s\Big).
\end{aligned}$$

It is straightforward to show that if $W_N > q\frac{h}{h+c}sL$, then $y_1 > 0$and $y_2 > 0$, which would also imply $\det A > 0$.

Suppose, to obtain a contradiction, that $W_N \leqslant q\frac{h}{h+c}sL$ in an interior equilibrium. If $\det A = 0$ in this equilibrium, then the set of solutions to the system above is a subspace of dimension one or higher, and in particular there are at least two points that lie on the boundary of the parallelepiped bounded by $0 \leqslant \beta_{N1}, \beta_{N2} \leqslant W_N$, $0 \leqslant \beta_{S1}, \beta_{S2} \leqslant W_S$. Both of these points would be equilibria; however, by the argument above, only one point on the boundary may be an equilibrium. This contradiction proves that $\det A \neq 0$.

Since $\det A \neq 0$, the solution to the system may be found using Kramer's rule. Applying this rule and simplifying, we may find the share of sick individuals who stay at home, $\frac{g}{W_S}\beta_{S1} + \frac{1-g}{W_S}\beta_{S2}$. This quantity may be written as

$$\frac{1}{\det A} \times \frac{(c+h)W_N - hqsL}{(c+h)^2 W_N^2 W_S} Mhq\left(Mcfhr\left(1-t^2\right) + W_N(1-2g(1-g)(1-t))(c+h)\right).$$

In an interior equilibrium, this is positive, and therefore, since $W_N \leqslant q\frac{h}{h+c}sL$, we must have $W_N < q\frac{h}{h+c}sL$ and $\det A < 0$. Let us now increase W_N. For some value $W_N^* < q\frac{h}{h+c}sL$, the value of $\det A$ would become zero. This means that for some lower $W_N' \in (W_N, W_N^*)$, the solution to the system will be on the boundary of the parallelepiped. As shown above, however, in any such equilibrium all sick individuals go out, which would not be the case for W_N' as $W_N' < W_N^* < q\frac{h}{h+c}sL$. This contradiction shows that in an interior equilibrium, we must have $W_N > q\frac{h}{h+c}sL$, and thus $\det A > 0$.

Now, if $W_N > q\frac{h}{h+c}sL$, it is straightforward to find the equilibrium cutoffs by applying Kramer's rule and show that the solution is indeed interior. It remains to show that $\beta_{N1} < \beta_{N2}$ and $\beta_{S1} < \beta_{S2}$. Suppose, to obtain a contradiction, that $\beta_{N1} \geqslant \beta_{N2}$. This implies $\gamma_{N1} \leqslant \gamma_{N2}$, and therefore $g_1\gamma_{N1} + g_2 t\gamma_{N2} < g_1 t\gamma_{N1} + g_2\gamma_{N2}$, as $t < 1$ and $g_1 < g_2$. From the first two equations in the system, we have $\beta_{N1} < \beta_{N2}$, a contradiction. The fact that $\beta_{S1} < \beta_{S2}$ may be proven similarly. This completes the proof. □.

Proof of Proposition C2. The proof involves using Kramer's rule to solve for equilibrium cutoffs and differentiating. For example, let us compute the comparative statics of the total share of people who self-isolate,

$$\sigma = (c+h)g\frac{\beta_{N1}}{W_N} + (c+h)(1-g)\frac{\beta_{N2}}{W_N} + sg\frac{\beta_{S1}}{W_S} + s(1-g)\frac{\beta_{S2}}{W_S}.$$

Using notation from the previous proof, it can then be written as

$$\sigma = \frac{x_1 g(1-g) + x_2}{y_1 g(1-g) + y_2} = \frac{x_1 v + x_2}{y_1 v + y_2},$$

where x_1, x_2, y_1, y_2 do not depend on g (we do not spell out x_1 and x_2 to save on space) and where we denoted $v = g(1-g)$. Since $g < \frac{1}{2}$, v is monotonically increasing in g. Consequently, the sign of $\frac{d\sigma}{dv}$ is the same as the sign of $x_1 y_2 - x_2 y_1$. We can show that

$$x_1y_2 - x_2y_1 = MhW_NW_S(1-t)(c+h)(2(c+h)W_N + Mchr(1-t))((c+h)W_N - Lhqs) \times \left(-c^2rW_S - chrW_S - cqsW_N - hqsW_N + Lh^2q^2s + Lchq^2s - Lchqrs\right).$$

Here, on the first line, all terms are positive (the last one is so because the equilibrium is unique). The term on the second line may be written as

$$qs(h+c)(qhL - W_N) - c((h+c)W_S + qhsL)r.$$

This is decreasing in r, and is positive if and only if $r < \frac{qs(h+c)(hqL-W_N)}{c((h+c)W_S+hqsL)} = \frac{qs}{c}\frac{qhL-W_N}{\left(W_S+q\frac{h}{h+c}sL\right)}$. Notice that

$$\frac{d}{dh}\frac{qs}{c}\frac{qhL - W_N}{\left(W_S + q\frac{h}{h+c}sL\right)} = \frac{q^2s}{c}\frac{(h+c)^2W_S + csW_N + qh^2sL}{((h+c)W_S + qhsL)^2}L > 0,$$

which means that the condition for $\frac{d\sigma}{dv}$ to be positive is more likely to be satisfied for higher h. Furthermore, in the light of Assumption C1, for h close to 1 and r close to 0, it is positive. The other results are proven in a similar way.

This completes the proof. □.

The proofs of Proposition C3 and C4, as well as the proof of Proposition 1, follow directly from these results.

Appendix D. Empirical results for the United States

To provide evidence that the results reported in the previous section are not specific to Russia, we replicate the analysis using data from the United States. Specifically, we use county-level data from SafeGraph on mobile devices to see whether the results of estimation of Eq. (1) are consistent with the ones that we get based on Russian data.

Background COVID-19 in the United States began showing up almost simultaneously in several states. By March 1, it had spread to nine states, including the first three hot spots, California, New York, and Washington. New York soon became the hardest-hit state. However, due to the geographically diverse early spread, the predictive power of interstate migration patterns with New York on the initial COVID-19 spread is much lower compared to the case of Moscow and Russia. For this reason, we are not able to use the IV approach in the U.S. setting and have to rely on the standard difference-in-differences estimation only.

Ethnic fractionalization is also a highly relevant topic for the United States, which has one of the most ethnically diverse populations in the world. Typically, however, in the American context, instead of ethnicities, diversity is discussed in terms of race. As of the 2018 Census Bureau's American Community Survey, 60% of the U.S. population was White, 18% Hispanic, 12% African American, and 6% Asian.[28] Still, states and counties vary drastically in their levels of ethnic (and racial) diversity: on one extreme, 93% of Maine's population is White, whereas Texas is split roughly 40%–40%–12%–5% among Whites, Hispanics, African Americans, and Asians. For many historical reasons, however, the U.S. population is highly segregated, and ethnic fractionalization correlates with many county-level characteristics, including population density. For this reason, in our analysis, we do our best to control for various confounders of ethnic diversity.

Data. To measure social distancing in the United States, we use the social distancing metrics compiled and released by SafeGraph.[29] The data are generated using a panel of GPS pings from anonymous mobile devices. Similar to much of the literature (e.g., Chiou and Tucker, 2020; Kapoor et al., 2020), we use the share of devices remaining completely home on a given day in a given county as the dependent variable. For each device, "home" location is determined by SafeGraph as the common nighttime location of each mobile device over a six-week period. Since the data are presented at the census-block level, we aggregate them up to the county level by taking the sum of all devices and the sum of all devices remaining completely home. We then calculate the county-level daily share by dividing the latter number by the former.

We use data on COVID-19 cases and deaths over time from the *New York Times* open repository on coronavirus cases.[30] From this source, we obtain data on the daily number of cases and deaths in each county and state. We accessed these data on May 5, 2020.

We draw data on counties' ethnic compositions from the 2010 Census, based on which we calculate the standard measure of ethnic (racial) fractionalization. For other county-level controls, such as population density, median household income, and the share of adults with a BA degree, we rely on the county-level benchmark

[28] https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/

[29] SafeGraph is a data company that aggregates anonymized location data from numerous applications to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group. For details on this particular dataset, see: https://docs.safegraph.com/docs/social-distancing-metrics.

[30] https://github.com/nytimes/covid-19-data.

indicators from the Social Capital Project.[31] Finally, we obtain data on state-level stay-at-home measures from Raifman et al. (2020). Table B.2 presents the summary statistics of all variables used in our analysis of the U.S. case.

Empirical Results. Table D.1 presents the results of the difference-in-differences estimation, similar to Table 1 for the Russian case.

The results are largely consistent with the Russian case. The magnitudes imply that following the discovery of the first case, the share of those staying at home increased by 1.9 percentage points on average for the most fractionalized county compared to the least fractionalized county. In other words, the difference between the counties with the highest and lowest fractionalization can explain 6.1% of average mobility reduction after the discovery of the first case or, alternatively, 8.2% of weekday-weekend gap for an average county.

Similarly to Table B.4, we report regressions that include interaction terms both with the report of the first case in the state and with the state-level stay-at-home orders. We summarize these results in Table D.2. We find, similar to the Russian case, no differential effect of statewide stay-at-home orders on the likelihood of staying at home depending on the level of ethnic fractionalization. At the same time, even in this demanding specification, the coefficient for the interaction between the dummy for the first reported case and ethnic fractionalization remains positive and significant in three out of four specifications.

Appendix E. Implications for mortality

While it is important to document the differential mobility reduction by ethnic fractionalization on its own, it is also of interest to see the implications of this differential effect for the spread of the disease. To this end, we produce some back-of-the-envelope estimates of how many premature deaths may have been prevented by greater social distancing in more diverse communities. Because the elasticity of deaths with respect to social distancing is unknown at this point, we rely on two estimates—one from a widely cited epidemiological study, and one based on the local average treatment effect estimated in the economic literature.

Elasticity of COVID-19 deaths with respect to social distancing. Based on an epidemiological model, Walker et al. (2020) predict that a uniform 45% reduction in interpersonal contact rate within a country would lead to a 50% reduction in mortality rate in Europe and North America, from eight deaths per 1,000 people to four deaths per 1,000 people.

In economics, Kapoor et al. (2020) use variation in rainfall the weekend prior to the official government lockdown to produce the IV estimates of the effect of lower share of home-stayers on cases and deaths from COVID-19. According to Kapoor et al. (2020, p. 7), "a one percentage point increase in the number of people leaving home on the weekend before the shutdown causes case counts to rise by roughly 13 per 100,000, which translates to roughly one extra death per 100,000.".

One may think of the estimates from Walker et al. (2020) as the upper bound, because they assume a permanent reduction in interpersonal contact and consider the full counterfactual of exponential growth. In contrast, the numbers from Kapoor et al. (2020) need to be viewed as the lower bound, because they study a temporary reduction in social distancing on one particular weekend and because they consider only the data available at the time they wrote their article.

[31] Available at https://www.lee.senate.gov/public/index.cfm/scp-index.

Russia. First, we produce a back-of-the-envelope estimate of the lives potentially saved in Russia. We note that, according to our estimates in columns 2 and 4 in Table 3, a one-standard-deviation increase of ethnic fractionalization (0.172) is associated with a $0.29 \times 0.172 \approx 0.05$ increase in the isolation index after the first case in the region. We also note that the pre-first-case median in the isolation index is 1.4, which means that a 0.05 increase in the isolation index is associated with a 3.5% decline in social mobility.

For the upper-bound estimate based on the epidemiological literature, we assume that mobility reduction equates to reduction in interpersonal contact.[32] Furthermore, we assume that the estimates from Walker et al. (2020) can be applied linearly with the same ratio, i.e., that a 1% reduction in interpersonal contact is always associated with a 1.1% reduction in mortality rates. Then, under these assumptions, one finds that a 3.5% reduction in social contact is associated with a 3.85% reduction in mortality rates. For Europe and North America, a 3.85% reduction from 4 deaths per 1,000 population is 0.154 fewer deaths per 1,000 population (see Fig. B.3 in Walker et al. (2020)). For Russia, this equates to $0.154 \times 144,500 = 22,250$ fewer deaths.

For the lower-bound estimate, we assume that the a one-standard-deviation increase in isolation index in Russia is associated with a one-standard-deviation increase in the share of people staying at home in the United States, i.e., that they are both measuring the same underlying factor. Under this assumption, a 0.05 increase in the isolation index is equivalent to a $0.05 \times (6.63/0.85) = 0.39$-percentage-point increase in the share of people staying home. Under the assumption that the U.S. calculations in Kapoor et al. (2020) apply to Russia, this equates to $0.39 \times 1,450 = 565$ fewer COVID-19 deaths (out of around 5,971 as of June 8, 2020).[33]

United States. In the United States, according to our estimates in columns 2 and 4 in Table D.1, a one-standard-deviation increase in ethnic fractionalization (0.252) is associated with a $0.252 \times 2.5 = 0.63$-percentage-point increase in share of people staying home.

We start with the lower-bound estimate, which is straightforward to compute, given that the estimates in Kapoor et al. (2020) rely on the same data from SafeGraph and the same variable of the share of people staying home. Under the assumption that the effect observed in Kapoor et al. (2020) is a LATE that is applicable to our "compliers" and that it is stable over time, a 0.63-percentage-point increase in the share of people staying home is associated with 0.63 fewer deaths per 100,000 people. In the United States, it equates to $0.63 \times 3,282 = 2,000$ fewer deaths (out of around 113,000 as of June 8, 2020).

For the upper-bound estimate, we rely on the same assumption as earlier. Since the pre-first-case median in the share staying home is 22%, a 0.63-percentage-point increase in the share of people staying home equates to a 2.8% increase in social distancing, or, as we assume, a 2.8% reduction in interpersonal contact. Using the same 1.1 ratio as above, a 2.8% reduction in social contact is associated with a 3.08% reduction in mortality rates. For Europe and North America, a 3.08% reduction from 4 deaths per 1,000 people is 0.123 fewer deaths per 1,000 people. Thus, for the United States, this is equivalent to $0.123 \times 328,200 \approx 40,400$, or roughly 40,000 fewer deaths.

[32] In principle, this need not be the case, since one can move around and still adhere to strict social distancing rules.

[33] Note that while the upper-bound estimates above consider all potential future deaths from the disease, these lower-bound estimates are calculated assuming that no deaths would occur starting the day after the Kapoor et al. (2020)'s estimates were produced. This explains the gulf between the two estimates.

Table D.1
Social Distancing, First Case, and Ethnic Fractionalization, U.S. data.

	% Staying Home			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	3.615***	2.545**	3.599***	2.526**
	[1.280]	[1.198]	[1.295]	[1.225]
Post First Case x Education		0.115***		0.114***
		[0.026]		[0.026]
Post First Case x Median HH Income (in '000s)		0.134***		0.135***
		[0.027]		[0.027]
Post First Case x Population Density		0.220**		0.223**
		[0.091]		[0.092]
County Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Days	115	115	115	115
Counties	3,138	3,137	3,138	3,137
Observations	360,870	360,755	360,870	360,755
R-squared	0.719	0.742	0.786	0.809

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Standard errors in brackets are clustered at the state level. Percentage of people staying home is calculated based on the number of mobile devices never leaving the house on a given day divided by the number of mobile devices observed in the county that day. The period is January 1 to April 24, 2020. Post-first-case indicator is equal to one after a county's state had its first COVID-19 case, and zero otherwise.

Table D.2
Social Distancing, First Case, Stay-at-Home Orders, and Ethnic Fractionalization, U.S. data.

	% Staying Home			
VARIABLES	(1)	(2)	(3)	(4)
Post First Case x Ethnic Fractionalization	2.875***	2.138*	2.810***	2.072
	[0.876]	[1.241]	[0.899]	[1.257]
Post First Case	−1.350***	−9.950***	−1.247***	−9.876***
	[0.305]	[1.040]	[0.336]	[1.059]
Stay at Home Measures x Ethnic Fractionalization	1.630	0.980	1.764	1.110
	[2.000]	[1.928]	[1.949]	[1.881]
Stay at Home Measures	2.091***	1.960***	2.053***	1.918***
	[0.556]	[0.611]	[0.548]	[0.609]
Post First Case x % Education		0.106***		0.105***
		[0.026]		[0.027]
Post First Case x Median HH Income (in '000s)		0.135***		0.136***
		[0.026]		[0.027]
Post First Case x Population Density		0.202**		0.205**
		[0.087]		[0.088]
City Fixed Effects	Yes	Yes	Yes	Yes
Day of the Week and Calendar Week Fixed Effects	Yes	Yes		
Day Fixed Effects			Yes	Yes
Days	115	115	115	115
Counties	3,138	3,137	3,138	3,137
Observations	360,870	360,755	360,870	360,755
R-squared	0.724	0.745	0.791	0.812

Notes: *** p < 0.01, ** p < 0.05, * p < 0.1. Standard errors in brackets are clustered at the state level. Percentage of people staying home is calculated based on the number of mobile devices never leaving the house on a given day divided by the number of mobile devices observed in the county that day. The period is January 1 to April 24, 2020. Post-first-case indicator is equal to one after a county's state had its first COVID-19 case, and zero otherwise.

References

Ager, P., Brueckner, M., 2018. Immigrants' genes: genetic diversity and economic development in the United States. Econ. Inq. 56, 1149–1164.

Alesina, A., Baqir, R., Easterly, W., 1999. Public goods and ethnic divisions. Quart. J. Econ. 114, 1243–1284.

Alesina, A., Ferrara, E.L., 2005. Ethnic diversity and economic performance. J. Econ. Lit. 43, 762–800.

Alesina, A., Harnoss, J., Rapoport, H., 2016. Birthplace diversity and economic prosperity. J. Econ. Growth 21, 101–138.

Alesina, A., La Ferrara, E., 2000. Participation in heterogeneous communities*. Q. J. Econ. 115, 847–904.

Algan, Y., Hémet, C., Laitin, D.D., 2016. The social effects of ethnic diversity at the local level: a natural experiment with exogenous residential allocation. J. Polit. Econ. 124, 696–733.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., Yang, D.Y., 2020. Polarization and Public Health: Partisan Differences in Social Distancing during COVID-19. Available at SSRN 3570274.

Altonji, J., Card, D., 1991. The effects of immigration on the labor market outcomes of less-skilled natives. In: Immigration, Trade, and the Labor Market, National Bureau of Economic Research Inc, pp. 201–234.

Ananyev, M., Poyker, M., Tian, Y., 2020. The Safest Time to Fly: Pandemic Response in the Era of Fox News, Working Paper.

Andersen, M., 2020. Early evidence on social distancing in response to COVID-19 in the United States, Available at SSRN 3569368.

Arbatli, C.E., Ashraf, Q.H., Galor, O., Klemp, M., 2020. Diversity and conflict. Econometrica 88, 727–797.

Barrios, J.M., Benmelech, E., Hochberg, Y.V., Sapienza, P., Zingales, L., 2020. Civic capital and social distancing during the covid-19 pandemic, University of Chicago, Becker Friedman Institute for Economics Working Paper.

Barrios, J.M., Hochberg, Y.V., 2020. Risk perception through the lens of politics in the time of the COVID-19 pandemic, University of Chicago, Becker Friedman Institute for Economics Working Paper.

Bartos, V., Bauer, M., Cahlíková, J., Chytilová, J., 2020. Covid-19 Crisis Fuels Hostility against Foreigners, CESifo Working Paper.

Borgonovi, F. and E. Andrieu (2020): "Bowling together by bowling alone: Social capital and Covid-19," Working Paper.

Brzezinski, A., Deiana, G., Kecht, V., Van Dijcke, D., 2020. The COVID-19 Pandemic: Government vs. Community Action Across the United States, Working paper.

Bursztyn, L., Egorov, G., Enikolopov, R., Petrova, M., 2019. Social media and xenophobia: evidence from Russia, NBER Working Paper.

Bursztyn, L., Rao, A., Roth, C., Yanagizawa-Drott, D., 2020. Misinformation During a Pandemic, Working paper.
Campante, F.R., Depetris-Chauvin, E., Durante, R., 2020. The Virus of Fear: The Political Impact of Ebola in the US, NBER Working Paper.
Card, D., 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. J. Lab. Econ. 19, 22–64.
Chetty, R., Hendren, N., Jones, M.R., Porter, S.R., 2019. Race and economic opportunity in the united states: an intergenerational perspective. Q. J. Econ. 135, 711–783.
Chiou, L., Tucker, C., 2020. Social Distancing, Internet Access and Inequality, NBER Working Paper.
Desmet, K., Wacziarg, R., 2020. Understanding Spatial Variation in COVID-19 across the United States, NBER Working Paper.
Doganoglu, T., Ozdenoren, E., 2020. Should I stay or should I go (out): The role of trust and norms in disease prevention during pandemics, Working Paper.
Durante, R., Guiso, L., Gulino, G., 2020. Asocial Capital: Civic Culture and Social Distancing During Covid-19, CEPR Discussion Paper 14820.
Easterly, W., Levine, R., 1997. Africa's growth tragedy: policies and ethnic divisions. Quart. J. Econ. 112, 1203–1250.
Ellickson, R.C., 1994. Order Without Law. Harvard University Press.
Engle, S., Stromme, J., Zhou, A., 2020. Staying at Home: Mobility Effects of Covid-19, Available at SSRN.
Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. J. Econ. Perspect. 14, 159–181.
Gitmez, A., Sonin, K., Wright, A.L., 2020. Political Economy of Shelter-in-Place Compliance, Working Paper.
Goette, L., Huffman, D., Meier, S., 2006. The impact of group membership on cooperation and norm enforcement: evidence using random assignment to real social groups. Am. Econ. Rev. 96, 212–216.
Guriev, S., Melnikov, N., 2016. War, inflation, and social capital. Am. Econ. Rev. 106, 230–235.
Hjort, J., 2014. Ethnic divisions and production in firms. Q. J. Econ. 129, 1899–1946.
Kapoor, R., Rho, H., Sangha, K., Bhavyaa, S., Ajay, S., Xu, G., 2020. God is in the Rain: The Impact of Rainfall-Induced Early Social Distancing on COVID-19 Outbreaks, Working paper.
Korovkin, V., Makarin, A., 2019. Conflict and Inter-Group Trade: Evidence from the 2014 Russia-Ukraine Crisis. Available at SSRN 3397276.
Lee, N., 2015. Migrant and ethnic diversity, cities and innovation: Firm effects or city effects? J. Econ. Geogr. 15, 769–796.
Miguel, E., Gugerty, M.K., 2005. Ethnic diversity, social sanctions, and public goods in Kenya. J. Public Econ. 89, 2325–2368.
Mikhailova, T., Valsecchi, M., 2020. Internal migration and Covid-19 (in Russian). In: Economic Policy in Times of Covid-19, New Economic School, pp. 26–33.
Montalvo, J.G., Reynal-Querol, M., 2005. Ethnic polarization, potential conflict, and civil wars. Am. Econ. Rev. 95, 796–816.
Montalvo, J.G., Reynal-Querol, M., 2020. Ethnic diversity and growth: Revisiting the evidence. Rev. Econ. Stat. forthcoming.
Ostrom, E., 1990. Governing the commons: The evolution of institutions for collective action. Cambridge University Press.
Ottaviano, G.I., Peri, G., 2006. The economic value of cultural diversity: evidence from US cities. J. Econ. Geogr. 6, 9–44.
Painter, M., Qiu, T., 2020. Political Beliefs affect Compliance with COVID-19 Social Distancing Orders, Available at SSRN 3569098.
Peri, G., 2012. The effect of immigration on productivity: Evidence from US states. Rev. Econ. Stat. 94, 348–358.
Putnam, R.D., 2007. E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture. Scand. Polit. Stud. 30, 137–174.
Raifman, J., Nocka, K., Jones, D., Bor, J., Lipson, S., Jay, J., Chan, P., 2020. COVID-19 US state policy database. Available at: www.tinyurl.com/statepolicies.
Rohner, D., Thoenig, M., Zilibotti, F., 2013. War signals: a theory of trade, trust, and conflict. Rev. Econ. Stud. 80, 1114–1147.
Ross, C.T., 2015. A multi-level Bayesian analysis of racial bias in police shootings at the county-level in the United States. PLoS ONE 10 (11).
Simonov, A., Sacher, S., Dubé, J.-P., Biswas, S., 2020. The Persuasive Effect of Fox News: Non-Compliance with Social Distancing During the Covid-19 Pandemic, Working Paper.
Stephens-Davidowitz, S., 2014. The cost of racial animus on a black candidate: Evidence using Google search data. J. Public Econ. 118, 26–40.
Valsecchi, M., Durante, R., 2020. Internal migration and the spread of COVID-19, Working Paper.
Walker, P.G., Whittaker, C., Watson, O., Baguelin, M., Ainslie, K., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Cattarino, L., et al., 2020. The global impact of COVID-19 and strategies for mitigation and suppression, Imperial College London.
Wright, A.L., Sonin, K., Driscoll, J., Wilson, J., 2020. Poverty and Economic Dislocation Reduce Compliance with COVID-19 Shelter-in-Place Protocols, University of Chicago, Becker Friedman Institute for Economics Working Paper.
Xiao, K., 2020. Saving Lives versus Saving Livelihoods: Can Big Data Technology Solve the Pandemic Dilemma? Working paper.