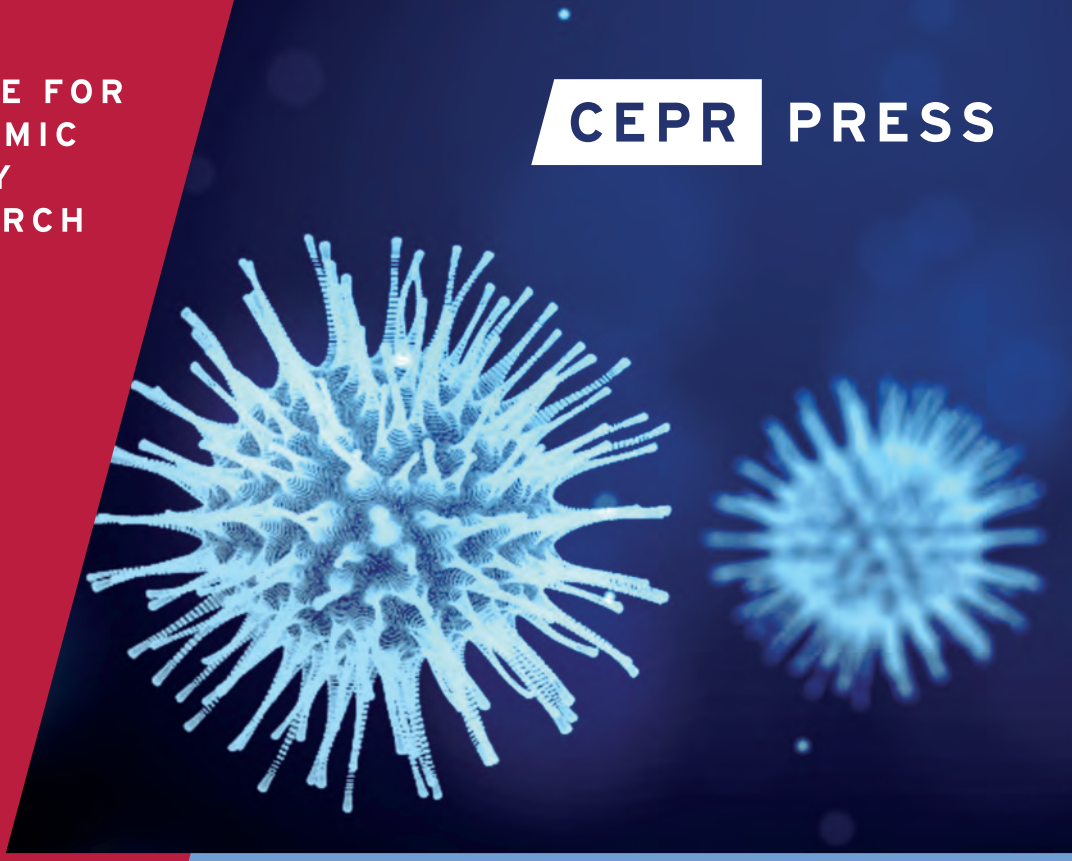
CENTRE FOR ECONOMIC POLICY RESEARCH

CEPR PRESS



COVID ECONOMICS

VETTED AND REAL-TIME PAPERS

ISSUE 53
23 OCTOBER 2020

Covid Economics
Vetted and Real-Time Papers

Covid Economics, Vetted and Real-Time Papers, from CEPR, brings together formal investigations on the economic issues emanating from the Covid outbreak, based on explicit theory and/or empirical evidence, to improve the knowledge base.

Founder: Beatrice Weder di Mauro, President of CEPR
Editor: Charles Wyplosz, Graduate Institute Geneva and CEPR

Contact: Submissions should be made at https://portal.cepr.org/call-papers-covid-economics. Other queries should be sent to covidecon@cepr.org.

The Centre for Economic Policy Research (CEPR)

The Centre for Economic Policy Research (CEPR) is a network of over 1,500 research economists based mostly in European universities. The Centre's goal is twofold: to promote world-class research, and to get the policy-relevant results into the hands of key decision-makers. CEPR's guiding principle is 'Research excellence with policy relevance'. A registered charity since it was founded in 1983, CEPR is independent of all public and private interest groups. It takes no institutional stand on economic policy matters and its core funding comes from its Institutional Members and sales of publications. Because it draws on such a large network of researchers, its output reflects a broad spectrum of individual viewpoints as well as perspectives drawn from civil society. CEPR research may include views on policy, but the Trustees of the Centre do not give prior review to its publications. The opinions expressed in this report are those of the authors and not those of CEPR.

Editorial Board

Ethics

Covid Economics will feature high quality analyses of economic aspects of the health crisis. However, the pandemic also raises a number of complex ethical issues. Economists tend to think about trade-offs, in this case lives vs. costs, patient selection at a time of scarcity, and more. In the spirit of academic freedom, neither the Editors of *Covid Economics* nor CEPR take a stand on these issues and therefore do not bear any responsibility for views expressed in the articles.

Submission to professional journals

The following journals have indicated that they will accept submissions of papers featured in *Covid Economics* because they are working papers. Most expect revised versions. This list will be updated regularly.

American Economic Review
American Economic Review, Applied Economics
American Economic Review, Insights
American Economic Review, Economic Policy
American Economic Review, Macroeconomics
American Economic Review, Microeconomics
American Journal of Health Economics
Canadian Journal of Economics
*Econometrica**
Economic Journal
Economics of Disasters and Climate Change
International Economic Review
Journal of Development Economics
*Journal of Econometrics**
Journal of Economic Growth
Journal of Economic Theory
*Journal of the European Economic Association**
Journal of Finance
Journal of Financial Economics
Journal of International Economics
*Journal of Labor Economics**
Journal of Monetary Economics
Journal of Public Economics
Journal of Public Finance and Public Choice
Journal of Political Economy
Journal of Population Economics
Quarterly Journal of Economics
*Review of Corporate Finance Studies**
Review of Economics and Statistics
*Review of Economic Studies**
Review of Financial Studies

(*) Must be a significantly revised and extended version of the paper featured in *Covid Economics*.

Covid Economics
Vetted and Real-Time Papers

Issue 53, 23 October 2020

Contents

Covid Economics Issue 53, 23 October 2020

Implications of heterogeneous SIR models for analyses of COVID-19[1]

Glenn Ellison[2]

Date submitted: 16 October 2020; Date accepted: 16 October 2020

This paper starts with a quick overview of results on the classic SIR model and variants allowing for heterogeneity in contact rates. It then notes several implications relevant to model calibrations and policy predictions. Calibrating the classic SIR model to data generated by a heterogeneous model can lead to forecasts that are biased in several ways and to understatement of the forecast uncertainty. Among the biases are that we may underestimate how quickly herd immunity might be reached, underestimate differences across regions, and have biased estimates of the impact of endogenous and policy-driven social distancing.

Covid Economics 53, 23 October 2020: 1-32

1 I thank Daron Acemoglu, Chris Avery, Victor Chernozhukov, Adam Clark, Jonathan Dushoff, Sara Fisher Ellison, Jim Stock, and Ivan Werning for helpful conversations and comments and Chris Ackerman and Bryan Kim for research assistance.

2 Gregory K. Palm Professor of Economics, Massachusetts Institute of Technology.

Covid Economics 53, 23 October 2020: 1-32

1 Introduction

The economic literature on the COVID-19 epidemic has developed at a remarkable pace. A number of recent economics papers build on the classic Susceptible-Infectious-Recovered (SIR) model to study how the epidemic may progress and how it may be affected by various policies.[1] In this note I review some results from the epidemiological literature on an SIR extension that economists have mostly not yet adopted, incorporating heterogeneity in the activity rates of different subpopulations, and note ways in which analyses based on classic SIR models can potentially yield misleading views.

The classic SIR model of Kermack and McKendrick (1927) has been a foundational model in epidemiology for nearly a century. It illustrates basic tradeoffs and provides a simple framework that can be easily built on. Subsequent work in epidemiological theory has extended the model in various ways, and modern epidemiological forecasts typically work with variants that are more flexible in a number of dimensions.[2] In this paper I focus on some theoretical extensions developed in the 1980's and 1990's that seem quite relevant to the COVID-19 epidemic. Specifically, I discuss two classic models that focus on heterogeneity in the frequency with which different individuals engage in interactions that risk spreading the disease. Given the current understanding about how COVID-19 seems to be transmitted, it is easy to think of a number of subpopulations who will have many more risky interactions than average: those living in overcrowded urban apartments, frequenting bars and nightclubs, using public transportation, attending crowded religious services, working in a nursing home, etc. Others, e.g. farmers and those who are retired or work from home, should be relatively safe.

Section 2 reviews of the classic SIR model and extensions. Each extension discussed is a multipopulation SIR model that supposes that the subpopulations differ in their "activity" levels. As with the classic SIR model, the differential equations describing the rates at which members of each subpopulation transition from the susceptible to the infectious state can be motivated by a process in which agents are randomly matched in continuous time with each interaction between susceptible and infectious agents potentially leading to a new infection. One version assumes "uniform" matching in which the probability that any two agents are randomly matched is proportional to the product of their activity levels. The other assumes "homophilic" matching in which agents are more likely

[1] See among others Acemoglu et al. (2020), Alvarez, Argente, and Lippi (2020), Baqaee et al. (2020), Eichenbaum, Rebelo, and Trabandt (2020), Farboodi, Jarosch, and Shimer (2020), Fernández-Villaverde and Jones (2020), Jones, Philippon, and Venkateswaran (2020), and Rowthorn and Toxvaerd (2012)

[2] See, for example, Champredon et al. (2018), Unwin et al. (2020), and Viboud et al. (2018).

Covid Economics 53, 23 October 2020: 1-32

to interact with others in their own subpopulation.[3] Each model behaves much like the classic SIR model. Small infections initially grow at an exponential rate if a composite parameter analogous to R_0 is greater than one and new infections slow (and eventually die out) once the fraction with acquired immunity passes a "herd immunity" threshold. The composite R_0 and the herd immunity thresholds depend on the characteristics of the various subpopulations, and I review results from prior work to illustrate important principles about how epidemics spread in heterogeneous populations.

Sections 3 and 4 then draw out implications of these models for our analyses of COVID-19. Section 3 emphasizes that thinking about heterogeneity in contact patterns suggests that making predictions about the course of the COVID-19 epidemic and the impacts of reopening policies is inherently difficult. Heterogeneous models have more parameters that need to be calibrated. Long run outcomes can be sensitive to activity levels of the less active, and it is difficult to calibrate these parameters early in an epidemic when there are few cases in less-active communities. This is particularly true when one contemplates removing restrictions and thereby increasing activity among the currently inactive. Predictions based on classic SIR models that do not allow for heterogeneity may be overconfident.

Section 4 then focuses on ways in which conclusions drawn from applying homogeneous SIR models to a world that may be like a heterogeneous SIR model can be misleading. One observation, also found in Gomes et al. (2020) and Britton, Ball, and Trapman (2020), is that homogeneous SIR models may substantially overstate the fraction of the population that must be infected in order to achieve herd immunity. Intuitively, if a small high-contact group plays a central role in spreading the disease, then incidence will much higher in this group, and once many in this group have acquired immunity the epidemic may die out. Less obvious but still relevant effects are present with less extreme heterogeneity. A second related observation is that (targeted) lockdown polices can also be more cost effective in heterogeneous populations. There can be substantial gains either from taking permanent measures to reduce spread among the highly active or from temporarily locking down less active groups to minimize overshooting of herd immunity thresholds. The differences in dynamics also imply that time-series estimates of policy impacts may be biased. In each case, effects depend both on the magnitude of the heterogeneity that it present and on the degree of homophily in matching. The discussions attempt to bring out comparative statics and plausible magnitudes of effects.

The final section of the paper discusses some practical implications of the results. The message that we may be missing information for assessing reopening plans is troublesome when reopening is already

[3] Jackson and Lopez-Pintado (2013) discusses impacts of homophily and heterogeneity in a class of models that includes SIS models in which recovered agents are again susceptible.

Covid Economics 53, 23 October 2020: 1-32

upon us. But the models suggest fairly easy ways in which economists could extend their models and also point to data opportunities that might reduce the critical uncertainties. The messages that controlling the epidemic may not be as hard as it appears in some models and that herd immunity might not be as far off might be naïvely predicted said also give room for optimism.

This paper is related to a number of others in epidemiology and economics. The discussion of heteogeneous SIR models is a review of a literature in epidemiology that dates back to the late 1980s and mid 1990s, with the particular formulations drawing heavily on Dushoff and Levin (1995). Empirical epidemiologists have also for quite some time been interested in multipopulation SIR models both examine to interactions between age groups (important for other reasons for childhood diseases) and groups, e.g. health care workers in the Ebola epidemic, who play an important role in transmission.[4] Two very recent working papers in epidemiology have made observations similar to the observation in section 4.1 that herd immunity thresholds can be substantially lower in heterogeneous SIR models than in homogeneous SIR models. Gomes et al. (2020) graphs the herd immunity threshold as a function of the coefficient of variatiion in contact rates in a heterogeneous SEIR model, noting estimates of the coefficients of variation that have been previously reported for other diseases. Britton, Ball, and Trapman (2020) give herd immunity thresholds for an 18-group model calibrated to estimated interactions across 6 age groups with assumed low-activity and high-activity individuals assumed to have activity levels that are half and twice the average activity levels and discuss partial lockdown policies that hold the infection to this level. Recent papers in epidemiology are also broadly related in that their observations motivate examining heterogeneous transmission. Worobey et al. (2020) concludes that early imported cases formerly thought to have triggered epidemics in Washington and Italy appear to not be related to the subsequent epidemics there, suggesting that the communities in which they occurred had low enough R_0 so that the epidemics they started died out. Miller et al. (2020) examine transmission in Israel using full genome sequence and conclude that there are "high levels of transmission heterogeneity ... with between 1-10% of infected individuals resulting in 80% of secondary infections.

As noted earlier, the primary motivation for the paper is the large recent literature in economics that builds on SIR models. In this literature, Avery et al. (2020) informally discuss the potential relevance of transmission heterogeneity. Most closely related are several very recent papers, including Acemoglu et al. (2020), Baqaee et al. (2020), Favero, Ichino, and Rustichini (2020), and Rampini

[4]See, for example, Britton (1998), Champredon et al. (2018), Demiris and O'Neill (2005), and Lloyd-Smith et al. (2005).

Covid Economics 53, 23 October 2020: 1-32

(2020), that use calibrated multipopulation SIR models to examine the impact of COVID-19 mitigation policies, and in the case of Acemoglu et al. (2020) to identify optimal policies from a broad class. These papers use age-defined group structures to illustrate the substantial gains from age-targeted policies due to how dramatically death rates vary with age. They do not focus on the impact of contact heterogeneity, nor do most of the calibrations include within-age-group heterogeneity, which is presumably much larger than cross-age group heterogeneity, but three of them do include some heterogeneity in contact rates. Baqaee et al. (2020) calibrate a five by five matrix of age group to age group contact rates using both general contact survey data and a workplace proximity survey to reflect differences in occupational mixes across age groups. Acemoglu et al. (2020) use uniform mixing in their main analyses, but also calibrate a three by three age group contact matrix to data from another contact survey. The groups in Favero, Ichino, and Rustichini (2020) are age $\times$ activity based with medium- and high-activity individuals assumed to be 12% and 18% more active than the low-activity group.

2 Heterogeneous SIR Models

In this section I'll quickly review the standard SIR model and then spend more time on two heterogeneous versions drawing on previous results.

2.1 The standard homogeneous SIR model

A number of recent economic analyses of the COVID-19 epidemic build on a standard homogeneous SIR model.

Consider a population of unit mass. Assume that at each time t each member of the population is in one of three states: Susceptible, Infectious, or Recovered. Write $S(t)$, $I(t)$, and $R(t)$ for the fractions in each state at time t. Assume that the dynamics of these fractions are:

$$\begin{aligned}\dot{I}(t) &= S(t)I(t)R_0\gamma - \gamma I(t)\\ \dot{R}(t) &= \gamma I(t)\\ \dot{S}(t) &= -S(t)I(t)R_0\gamma\end{aligned}$$

One way to motivate the model is to suppose that agents are being uniformly randomly matched in continuous time. Each agent meets another with probability $R_0\gamma dt$ in a dt time interval. A susceptible agent matched with an infectious agent becomes infectious. Agents transition from the Infectious state

Covid Economics 53, 23 October 2020: 1-32

to the Recovered state at Poisson rate γ. These transitions reflect both true recoveries and deaths from the disease.

The parameter R_0 can be thought of as the expected number of people that a newly infected person will directly infect in a population where everyone is susceptible. It is the critical determinant of the behavior of the model. Three important facts are:

1. If $R_0 > 1$, then the equilibrium $(S, I, R) = (1, 0, 0)$ is locally unstable. Adding a small number of infected agents leads to contagious growth in I. Equilibria with $I = 0$ are locally stable if $R_0 < 1$. A small infection dies out.

2. The model has a "herd immunity" threshold of $\overline{S} \equiv 1/R_0$. Any state $(S, 0, 0)$ with $S < \overline{S}$ is a stable steady state, so a small infection introduced into such a population will not spread. This does not, however, mean that epidemics will not infect more than a fraction $1 - \overline{S}$ of the population. When the herd immunity threshold is first reached we have $\dot{I}(\overline{S}, I, R) = 0$. This means that the infectious rate is (locally) constant with new infections occurring as fast as people are recovering. If the herd immunity threshold is reached at a point when I is large (which it typically is in models with R_0 large), then there can be substantial "overshooting" and many more than $1 - \overline{S}$ people can eventually be infected.

3. Define the growth rate of the infectious population by $g(t) = \frac{d}{dt} \log(I(t))$. Then, $g(t) = \gamma(R_0 S(t) - 1)$.

In the initial phase of an epidemic when $S(t) \approx 1$, the third fact says that the growth rate of the infectious population is approximately $\gamma(R_0 - 1)$. One can think of this as a cumulative growth rate of $R_0 - 1$ over the $1/\gamma$ average duration of an infection.

Investigations of whether restrictions are "flattening the curve" often graph the log of cumulative infections, i.e. $\log(1 - S(t))$, versus time. This curve will be approximately linear with slope $\gamma(R_0 - 1)$ as long as the ever-infected fraction of the population remains small, e.g. when the US has had 10 million cases. Attempts to infer R_0 from such curves are common given the desire to assess where the herd immunity threshold might be.

Epidemiologists commonly work with extensions of the SIR model. Among the standard additions are an additional state E of agents who are infected but not yet infectious, more flexible recovery processes that allow non-exponential infectious durations, an explicit death state, and population

Covid Economics 53, 23 October 2020: 1-32

inflows/outflows. Some economic models also incorporate some of these elements. To simplify the discussion I will not incorporate any of these features here, but similar conclusions should apply.

2.2 A heterogeneous SIR model with uniform matching

In practice, some individuals are more interactive than others. For example, supermarket cashiers will be in the vicinity of many more people in a typical day than will retirees. Epidemiologists have also analyzed models that allow for such heterogeneity.[5]

A tractable version is motivated by uniform matching in a population consisting of N equally sized subpopulations indexed by $i = 1, 2, \ldots, N$. Suppose that members of group i are randomly matched with probability $R_{0i}\gamma dt$ in each dt time interval. Order the populations so that $R_{01} > R_{02} > \ldots > R_{0N}$. Assume that the matchings are uniform so that the probability that a matched agent from group i meets a group j agent is $R_{0j}/\sum_k R_{0k}$. Suppose any matching between a susceptible and an infectious agent results in the susceptible agent becoming infectious. Write $S_i(t)$, $I_i(t)$, and $R_i(t)$ for the fraction of agents in group i who are susceptible, infectious, and recovered at time t, and $S(t)$, $I(t)$, and $R(t)$ for the vectors with these terms as components.

With the same recovery process as before, this matching process motivates analyzing a system of differential equations:

$$\begin{aligned}
\dot{I}_i(t) &= S_i(t)\sum_j \beta_{ij} I_j(t) - \gamma I_i(t) \\
\dot{S}_i(t) &= -S_i(t)\sum_j \beta_{ij} I_j(t) \\
\dot{R}_i(t) &= \gamma I_i(t)
\end{aligned}$$

with $\beta_{ij} \equiv \gamma R_{0i}\frac{R_{0j}}{\sum_k R_{0k}}$.

With the assumption that the population size remains constant, the state is fully described by $S(t)$ and $I(t)$ and we will usually omit $R(t)$ from the state vector. For any vector S^0 giving the fraction of susceptibles in each group, the disease free state $(S, I) = (S^0, 0)$ is a steady state. To analyze the stability of such a steady state and the behavior of the system in a neighborhood thereof, we linearize the system around the steady state. Note also that all derivatives $\frac{\partial \dot{I}_i}{\partial S_j}$ are equal to zero when evaluated at a state with $I = 0$. Hence, the behavior of I in a neighborhood of $(S^0, 0)$ in the full 2N-dimensional

[5] See Andreasen and Christiansen (1989), Diekmann, Heesterbeek, and Metz (1990), Dushoff and Levin (1995), Hethcote (2000), Jacquez, Simon, and Koopman (1995), May and Anderson (1989), and Van den Driessche and Watmough (2002). The exposition below draws heavily on Dushoff and Levin (1995).

system has the same first-order approximation as that of I in the N-dimensional system

$$\dot{I} = A^{S^0} I,$$

where A^{S^0} is the partial derivative matrix with ijth element

$$a_{ij} = \left.\frac{\partial \dot{I}_i}{\partial I_j}\right|_{(S^0,0)} = \begin{cases} S_i^0 \beta_{ij} - \gamma & \text{if } j = i \\ S_i^0 \beta_{ij} & \text{if } j \neq i. \end{cases}$$

In particular, the equilibrium is locally stable if all eigenvalues of this matrix have negative real parts, and unstable if any eigenvalue has a positive real part.

The A^{S^0} matrix has positive off-diagonal elements, so the eigenvalue with the largest real part is real, and corresponds to a strictly positive eigenvector. This eigenvector gives the relative prevalence of the infected across groups for which the total number infected grows most rapidly. The special structure of this matrix allows one to easily find this eigenvector. It is $v_1 = (S_1^0 R_{01}, \ldots, S_N^0 R_{0N})$, i.e. prevalence is proportional to the product of the susceptible fraction and the contact rate. The eigenvalue corresponding to this eigenvector is

$$\lambda_1 = \gamma \left(\frac{\sum_i S_i^0 R_{0i}^2}{\sum_i R_{0i}} - 1 \right).$$

Two important implications of this are:

1. The equilibrium $(S^0, 0)$ is locally stable if $\frac{\sum_i S_i^0 R_{0i}^2}{\sum_i R_{0i}} < 1$ and locally unstable if $\frac{\sum_i S_i^0 R_{0i}^2}{\sum_i R_{0i}} > 1$.
2. For small δ, the growth rate of the log of the total infected population at the state $(S^0, \delta v_1)$ is approximately $\gamma \left(\frac{\sum_i S_i^0 R_{0i}^2}{\sum_i R_{0i}} - 1 \right)$.

Note that if we start from any state with a very small fraction δ infected, then the initial cases will initially grow at different rates in the different groups in a way that makes the distribution of cases across groups aligned with the principal eigenvector v_1.[6] Hence, provided that this alignment has already occurred by the time the epidemic starts to be measured, the early growth of a heterogeneous-SIR epidemic with activity vector R_0 will resemble the early growth of a homogeneous-SIR epidemic with parameter $\overline{R}_0 \equiv \frac{\sum_i R_{0i}^2}{\sum_i R_{0i}}$.

[6] Suppose the initial population infected is δv with $v = \sum a_i v_i$ where the v_i are the eigenvectors of A. In a neighborhood of this point we will have $I(t) \approx \sum a_i e^{\lambda_i t} v_i$, which becomes aligned with v_i.

Covid Economics 53, 23 October 2020: 1-32

Covid Economics 53, 23 October 2020: 1-32

Two ways of rewriting this expression are informative. First,

$$\overline{R}_0 = \frac{\sum_i R_{0i}^2}{\sum_i R_{0i}} = \sum_i \frac{R_{0i}}{\sum_k R_{0k}} R_{0i}.$$

This formula makes clear that growth rates depend on a weighted average of group-level R_{0i}'s, with the weights being proportional to the activity level in each group. This weighted average can be substantially higher than the unweighted mean. The relation to the unweighted average is made clearer by a second rewriting:

$$\overline{R}_0 = \frac{\sum_i R_{0i}^2}{\sum_i R_{0i}} = \frac{NE(R_{0i}^2)}{NE(R_{0i})} = E(R_{0i}) + \frac{\text{Var}(R_{0i})}{E(R_{0i})}.$$

This equality indicates that growth rate is the sum of the unweighted average of the R_{0i} and the ratio of the variance of the R_{0i} across groups to the mean. The latter can easily be quite important quantitatively.

2.3 A heterogeneous SIR model with homophily

While supermarket cashiers may interact with a fairly representative sample of the population, some other highly active groups disproportionately interact with others in their group. For example, those who frequent nightclubs, take public transportation, attend crowded religious services, or live in a working class neighborhood with overcrowded housing disproportionately interact with others who do the same things. Those who live in rural areas will disproportionately interact with others who live in the same rural area.

Heterogeneous SIR models with homophilic matching are more difficult to analyze, but epidemiologists have also derived insightful characterizations of some such models, referred to sometimes as models with "preferred mixing" or "like-with-like preference". To motivate one such model, consider an N group model as in the previous subsection, but suppose that when an agent from group i is randomly matched the probability that the person with whom they are matched is in group j is

$$p_{ij} = \begin{cases} h + (1-h)\frac{R_{0j}}{\sum_k R_{0k}} & \text{if } j = i \\ (1-h)\frac{R_{0j}}{\sum_k R_{0k}} & \text{if } j \neq i. \end{cases}$$

Such a matching process would lead to an SIR model nearly identical to that in the previous

subsection with

$$\dot{I}_i(t) = S_i(t) \sum_j \beta_{ij}^h I_j(t) - \gamma I_i(t),$$

where

$$\beta_{ij}^h = \begin{cases} \gamma R_{0i}(h + (1-h)\frac{R_{0j}}{\sum_k R_{0k}}) & \text{if } j = i \\ \gamma R_{0i}(1-h)\frac{R_{0j}}{\sum_k R_{0k}} & \text{if } j \neq i. \end{cases}$$

Once again, any $(S^0, 0)$ is a steady state of the system and we can analyze the stability of this steady state by looking at a linearized N-dimensional system:

$$\dot{I} = A^{S^0 h} I,$$

where $A^{S^0 h}$ is the partial derivative matrix with ijth element

$$a_{ij}^h = \left.\frac{\partial \dot{I}_i}{\partial I_j}\right|_{(S^0,0)} = \begin{cases} S_i^0 \beta_{ij}^h - \gamma & \text{if } j = i \\ S_i^0 \beta_{ij}^h & \text{if } j \neq i. \end{cases}$$

The off-diagonal elements of this matrix are again positive, so the eigenvector with the largest real part is again unique and corresponds to a positive eigenvector. It is no longer easy to give an explicit formula for the eigenvalue, but as noted by Diekmann, Heesterbeek, and Metz (1990) and Dushoff and Levin (1995) we can give explicit necessary and sufficient conditions for the equilibrium to be stable.

1. If $S_i^0 h R_{0i} > 1$ for any i, then $(S^0, 0)$ is unstable. This is obvious: the number of infected in population i will increase solely from within-group contacts, and cross-group contacts only add to the growth.

2. If $S_i^0 h R_{0i} < 1$ for all i, then $(S^0, 0)$ is unstable if $\sum_i \frac{R_{0i}}{\sum_k R_{0k}} \frac{1}{1 - h S_i^0 R_{0i}} (S_i^0 R_{0i} - 1) > 0$ and stable if $\sum_i \frac{R_{0i}}{\sum_k R_{0k}} \frac{1}{1 - h S_i^0 R_{0i}} (S_i^0 R_{0i} - 1) < 0$.

Note that when $h = 0$ the stability condition in the part 2. simplifies to a version of expression we gave earlier for the uniform model: $\sum_i \frac{R_{0i}}{\sum_k R_{0k}} (S_i^0 R_{0i} - 1) < 0$. For a disease-free equilibrium to be stable in a model with $h > 0$ it must satisfy the additional constraints in 1. that $S_i^0 h R_{0i} < 1$ for all i as well as the modified inequality given in 2. Note that the summation in this inequality differs from the summation for $h = 0$ in that we multiply the ith term by $\frac{1}{1 - h S^0 R_{0i}}$. These multiplicative factors are positive for all terms, and they are larger for the terms with $S_i^0 R_{0i}$ larger. Hence, we can think of the sum as proportional to a reweighting of the $h = 0$ sum that puts greater weight on the terms

Covid Economics 53, 23 October 2020: 1-32

with $S_i^0 R_{0i}$ large and less weight on the terms with $S_i^0 R_{0i}$ small. As a result, if the model with $h = 0$ is unstable, then the model with $h > 0$ is unstable as well.[7]

The argument above extends easily to a full monotonicity theorem: if a disease-free equilibrium is unstable for some value of h, then it is unstable for any value $h' > h$. Hence, a clear intuition one can take away from this model is that homophilic matching is an obstacle to the stability of disease free states.

Note that homophily on its own does not affect the dynamics of an epidemic. If we consider a model in which the R_{0i} are identical across groups, then as long as h is not extremely close to one, a small infection introduced into any one population will soon equalize across populations. With equal fractions infected in each population, the dynamics of the homophilic multipopulation model are identical to the $h = 0$ model. Hence, all of the effects of homophily discussed above should be understood as the effects of the combination of homophily and contact heterogeneity.

3 Challenges Inherent in Analyzing Heterogeneous Population Epidemics

In this section and the one that follows I turn to the task of drawing out implications of the above models for analyses of COVID-19. This section stresses a cautionary implication: it can be difficult to provide policy advice in epidemics that are well described by heterogeneous population SIR models. In particular, with currently available data it is challenging to estimate activity rates in less active populations, and important outcomes can be sensitive to these hard-to-estimate parameters.

3.1 Difficulty in calibrating models

Early in the COVID-19 epidemic several authors noted that it is difficult to calibrate critical parameters of homogeneous SIR model in the initial phase of an epidemic.[8] In the initial phase we may not have reliable data on anything but deaths. The fact that deaths in the model increase at an exponential rate makes it fairly easy to estimate R_0. But when many cases go unreported it is hard to calibrate the death rate. Different death rates would lead to dramatically different future paths of the epidemic. This weak identification problem goes away when we get some other piece of information that lets us estimate the death rate. Some potential sources for this are random serology tests to estimate the

[7] Mathematically, this is a classic Chebyshev inequality argument: if a_i and b_i are monotone increasing, then $\sum_i a_i b_i > \frac{1}{N} \sum_i a_i \sum_i b_i$.

[8] See Atkeson (2020), Fernández-Villaverde and Jones (2020), Korolev (2020), and Stock (2020).

Covid Economics 53, 23 October 2020: 1-32

Covid Economics 53, 23 October 2020: 1-32

fraction that have ever been infected, fatality data from locations, e.g. South Korea or the Diamond Princess, where we think almost all cases have been identified, or seeing an epidemic peak, which is informative about S.[9]

As more information has become available many economists have analyzed SIR models with calibrated R_0 and death rate parameters. The state of the art, in fact, has already moved well beyond this with a few recent working papers analyzing calibrated multipopulation SIR models that allow death rates and contacts to vary by age group.[10] The contact rate calibrations in these papers rely on three survey datasets. POLYMOD (Mossong et al. (2020)) and the BBC Pandemic Project (Klepac et al. (2020)) are survey datasets which asked respondents to list those with which they had contact in the previous 24 hours. And employment website O*Net asked workers in a large number of occupations to report how physically close to others they worked on a 5 point scale. One can also now capture some changes in activity over time using movement data available from firms with phone-tracking capabilities.

Heterogeneous SIR models that allow for idiosyncratic variation in contact rates by breaking a population (or each age group or other cell) into subpopulations that differ in activity levels have more parameters than do SIR models that do not consider such divisions. The fact that predictions can be dramatically affected by heterogeneity in R_0 suggests that it is important to try to capture some of the heterogeneity that surely exist within the cells that economists have been using with these extra parameters.

One approach to calibrating the extra parameters might be to use data on the variance of reported contacts in contact surveys. Although the surveys mentioned above have been used to estimate the relative prevalence of different age-group to age-group contacts, they seem less compelling as a source for estimating contact heterogeneity. For one thing, the way that contacts were defined, e.g. in the BBC survey contacts were defined as those whom one had physically touched or had a face-to-face conversation of at least three words with, leaves out many contacts that may be important in spreading COVID-19: singing near someone in a choir practice, standing near someone in a crowded bar, riding on the same subway train, being served by a cruise ship waiter, etc. The obvious heterogeneities in the frequency of such unrecorded contacts may mostly cancel out when one computes means for a large group, but we would definitely want to capture them to calibrate a model of contact hetero-

[9] Fernández-Villaverde and Jones (2020) note that more complex SIR models fit under a variety of assumptions about accessory parameters make very similar predictions about the future course of epidemics in locations where epidemics have peaked.

[10] See Acemoglu et al. (2020), Baqaee et al. (2020), and Favero, Ichino, and Rustichini (2020).

Covid Economics 53, 23 October 2020: 1-32

geneity. Another limitation of the main contact surveys is that they record contacts on a single day. Hence, recorded cross-subject variation confounds differences is cross-sectional means and time-series variation.

Another approach might be to calibrate the model to the path of the epidemic to-date. The initial growth rate of an epidemic should let us estimate the parameter composite $\overline{R}_0$ mentioned earlier. However, in a heterogeneous population SIR model, there is a weak identification problem when one tries to get more than this: it can be very difficult to obtain estimates of the activity rates in the less-active populations even after there has been substantial spread of the infection. Intuitively, when there is substantial heterogeneity in the R_{0i}, there will be a substantial number of infections when the epidemic surges in the highest R_{0i} subpopulations. At that point, there may still be few infections in many of the less-active groups, particularly if matching is homophilic. This can make it very difficult to estimate activity parameters for the low infection groups from aggregate infection data.

While it is hard to have any confidence in a calibration, I know that many economist readers will want to know if the differences between heterogeneous and homogeneous SIR models are salient for plausible parameters. Accordingly, I will at times discuss a simple numeric example in which the mean and variance of activity levels across groups has been chosen to be in the plausible range. Specifically, I will sometimes discuss a population with five equally-sized subpopulations having activity rates 3.5, 1.5, 1, 0.5, 0.5. With uniform matching the model has $\overline{R_0} \approx 2.3$ which roughly matches the the growth rates assumed by Ferguson et al. (2020) and Acemoglu et al. (2020).[11] The coefficient of variation of the cross-group differences, 0.8, roughly matches the variation in reported contacts in the BBC Pandemic Project data.

3.2 Difficulty in predicting future epidemic paths

The fact that some parameters of the heterogeneous SIR model are difficult to calibrate would not be troubling if the hard-to-estimate parameters of the model did not affect model predictions that we care about. Unfortunately, this is not true for the heterogeneous SIR model. One reason is that activity levels in the relatively low activity groups can have a substantial impact on the long run course of the epidemic. As an illustration, Figure 1 graphs new daily cases for two heterogeneous SIR models.[12]

[11] This is also consistent with some of the more sophisticated recent estimates growth rates such as that of Miller et al. (2020).

[12] Both models have ten equally-sized subpopulations with $h = 0.7$. The population with the long-lasting epidemic has $R_0 = (5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1, 1, 1)$. The population with the shorter-lived epidemic has $R_0 = (5.4, 2.6, 0.6, 0.4, \ldots, 0.4)$ and a lower fraction initially infected.

Covid Economics 53, 23 October 2020: 1-32

The parameters of the two models were chosen so that new cases take off at about the same time, rise to a peak at about the same rate, and peak at about the same level. Despite the nearly identical behavior up to the point when the epidemics peak, however, the epidemics proceed very differently on the way back down. In the end, one epidemic eventually infects more than twice as many people as the other, 58% vs. 28% of the population. The fraction who will eventually be infected under a given constant policy is obviously highly policy-relevant, and this example indicates that it will sometimes be very difficult to predict even when an epidemic is sufficiently far along as to have already reached its peak.

Intuitively, the way in which the example was constructed is that the two models each have fairly homophilic matching ($h = 0.7$) and feature a highly-active subpopulation in which the epidemic peaks before many in the less active subpopulations have been extensively infected. The models differ in the activity levels of the less-active. In one population, corresponding the dashed red line, seven of the ten subpopulations have $R_{0i} = 0.4$. The epidemic never really takes off in these groups and this results in the fairly rapid decline in infection rates once the epidemic has burned through the highly active groups. In the other population, corresponding to the solid blue simulation, nine of the ten groups have R_{0i} equal to 1.5 or 1.0. The infections coming out of the most active group set off a spread in these groups that goes on for quite some time. This produces an asymmetric peak with a decline that is much more gradual than the run up. Most of the total infections occur post-peak.

Analyses that do not consider the possibility that there may be heterogeneity in contact rates can report confidence intervals that are too narrow for this reason. For example, while the US remains quite far from the no-social-distancing herd immunity threshold, many states (and countries) are well beyond their peaks, which means that SIR-based models will make highly confident predictions about the future course of the epidemic presuming that individual behaviors and government policies remain fixed. For example, the April 29th update of the widely discussed IHME model gave its confidence interval for August 1st Massachusetts COVID-19 deaths as just 0 to 2.

3.3 Difficulty in predicting policy impacts

As reopening has become salient, a number of economic analyses have modeled the effects that relaxations of restrictions may have.[13] Thinking about heterogeneous models, however, suggests that it will be challenging to confidently make such predictions. Uncertainty about activity levels in low-activity

[13] See, for example, Baqaee et al. (2020).

CEPR PRESS

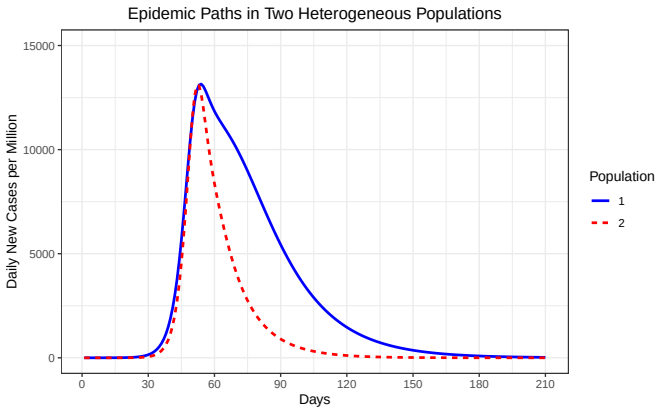


Figure 1: Example illustrating the difficulty in predicting the long-run course of a heterogeneous-population epidemic given the path of infection rates up to the point when the infection peaks. New daily cases are graphed for two ten population heterogeneous SIR models with $h = 0.7$. Model 1 has $R_0 = (5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1, 1, 1)$. Model 2 has $R_0 = (5.4, 2.6, 0.6, 0.4, \ldots, 0.4)$

populations can become even more important when considering policies relax social distancing. Intuitively, if we are far from the herd-immunity region (given the new policy), then the relaxation will set off a substantial second wave. If we are already close to or in the herd-immunity region, then the second wave will be smaller or nonexistent. Where we are relative to the herd immunity threshold depends on the full set of R_{0i}, including the hard-to-estimate activity levels in populations that have seen few infections while activity is tightly restricted.

Figure 2 provides a numerical illustration. It shows the time paths that an epidemic would follow under the same nonconstant policy path in two heterogeneous SIR populations. The policy involves a severe lockdown, reducing activity levels by 65%, imposed gradually over a two-week period just as the epidemic is taking off, and a partial relaxation about a month later that allows activity levels to return to 70% of their pre-lockdown values. The left panel plots new daily cases. The right panel plots cumulative cases to date. The vertical lines mark the dates when the initial lockdown starts its phase in and the date on which it is relaxed. The epidemics rise at very similar rates in the two populations prior to the lockdown. They have similar declines once the initial severe lockdown is imposed. Indeed, in the right panel it is very hard to see any difference in the courses of the two epidemics up through

Covid Economics 53, 23 October 2020: 1-32

Covid Economics 53, 23 October 2020: 1-32

the date at which the relaxation occurs.

Despite this similarity in the initial run up and through the lockdown, the two epidemics follow very different paths following the relaxation. As in the previous example, this reflects that the parameters were chosen so that activity levels in the less active subpopulations differ. In one population, whose outcomes correspond to the solid blue line, the relatively low activity populations have $R_{0i} = 1.5$. When we relax distancing rules, a large second wave takes off in these groups, infecting nearly three times as many people as had the first wave. In the other population, corresponding to the dotted red line, the low-activity populations have $R_{0i} = 0.7$ and this makes the second wave much smaller. Difficulty in distinguishing the blue from the red population at the point when the relaxation is occurring will make it difficult to predict which future course we should anticipate.

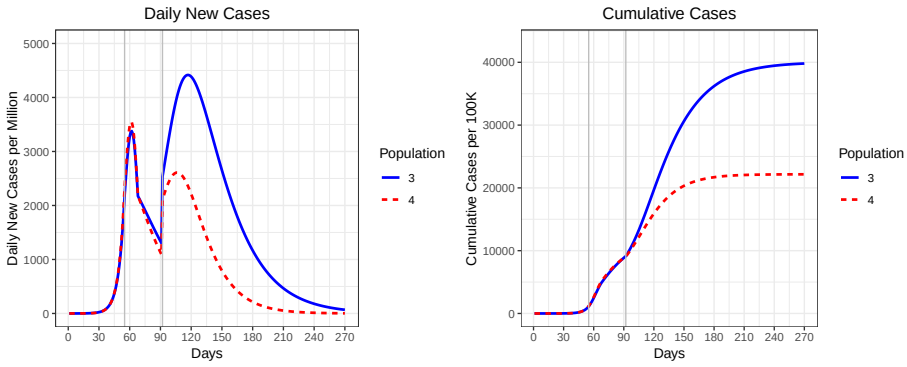


Figure 2: Example of epidemics that diverge after a policy relaxation. The figure graphs new daily cases and cumulative cases for heterogeneous SIR models with $h = 0.7$ under a policy intervention involving a severe lockdown and a partial relaxation. Model 3 has $R_0 = (3.63, 3.63, 1.5, \ldots, 1.5)$. Model 4 has $R_0 = (3.55, 3.55, 0.7, \ldots, 0.7)$.

4 Potential Biases From Ignoring Heterogeneity

Many economic analyses of the COVID-19 epidemic build on the simpler homogeneous SIR model, even though heterogeneous models seem more natural. This section notes several ways in which ignoring or understating heterogeneity in contact rates may bias the conclusions these analyses reach.

Covid Economics 53, 23 October 2020: 1-32

4.1 Overstatement of the damage incurred in reaching herd immunity

COVID-19 may remain widespread until after we pass a herd immunity threshold. Understanding how many cases must occur before herd immunity is reached is critical to assessing policies that lead to herd immunity.[14] An influential paper by Ferguson et al. (2020) suggested that deaths with uncontained spread could be very high and Greenstone and Nigam (2020) note that they correspond to extremely high economic costs under standard value-of-life assumptions. Two critiques of these calculations are that deaths may not be as high as the models suggest even without a government response due to endogenous social distancing, and that fatality rates could be lower due to asymptomatic cases. We note here another reason: models with heterogeneous activity suggest that herd immunity thresholds may be lower than naïve calculations based on homogeneous SIR models suggest.

In the homogeneous SIR model herd immunity is reached when $S = 1/R_0$, implying that the fraction of the population infected on the path to herd immunity must be at least $1 - 1/R_0$. If the system is instead described by the heterogeneous SIR model with uniform matching, then the naïve estimation of an R_0 parameter may lead us to misestimate the herd immunity threshold as

$$\hat{S} = \frac{1}{\overline{R}_0} = \frac{\sum_i R_{0i}}{\sum_i R_{0i}^2}.$$

This is indeed the threshold at which herd immunity is reached **if** the susceptible fraction is equal in all groups, but we can reach herd immunity with fewer infected by concentrating infections in the more active populations, and infections will naturally concentrate in the more active populations.

If $R_{0i} > 1$ for all i, then one state that obviously achieves herd immunity is to set $S_i = 1/R_{0i}$ for all i. The fraction susceptible is $\frac{1}{N}\sum_i 1/R_{0i}$. That this is always greater than $\hat{S}$ can be seen via an elegant two-step argument comparing both expressions to the reciprocal of the arithmetic mean of the R_{0i},

$$\frac{1}{N}\sum_i 1/R_{0i} = \frac{1}{1/\left(\frac{1}{N}\sum_i \frac{1}{R_{0i}}\right)} \geq \frac{1}{\frac{1}{N}\sum_i R_{0i}} \geq \frac{\frac{1}{N}\sum_i R_{0i}}{\frac{1}{N}\sum_i R_{0i}^2} = \hat{S},$$

with the two inequalities coming from the two parts of the root mean square-arithmetic-harmonic mean inequality. More important than the elegance is that the difference can be quite large in practical terms. For example, in our loosely calibrated five-population example with $R_0 = (3.5, 1.5, 1, 0.5, 0.5)$, the naïve homogeneous SIR calculation gives $\hat{S} = 7/16 \approx 0.44$, suggesting that 56% of the population

[14] Also critical to such calculations are an assessment of the extent to which we will overshoot herd immunity and the excess deaths that may occur due to exceeding hospital capacity.

must be infected before herd immunity is reached. However, the $S_i = 1/R_{0i}$ state is in the herd immunity region and has just 21% of the population infected.

More generally, the maximum fraction that can remain uninfected at the herd immunity point is calculated by solving

$$\max_{S_1,\ldots,S_N} \sum_i S_i$$

s.t.

$$\sum_i \frac{R_{0i}}{\sum_k R_{0k}} S_i R_{0i} \leq 1.$$

The linearity of the objective function and constraint make clear that the optimal solution involves concentrating the infections in the highest activity groups, i.e. S_i equal to zero in the highest-activity groups, S_i equal to one in the lowest activity groups, with S_i perhaps at an intermediate level in some marginal group to make the constraint hold with equality.[15] This will require even fewer infections than the $S_i = 1/R_{0i}$ state. In the five-population example above, we can achieve herd immunity with just 14% of the population infected by fully concentrating infections in the highest-activity subpopulation. While the example is clearly very loosely calibrated, the fact that the true level of infection needed to reach herd immunity is just one-fourth of what a naïve homogeneous-SIR based calculation indicates that contact heterogeneity is potentially a very important consideration.

For another example that may provide additional intuition, consider the spread of an epidemic in a less-developed country that lacks adequate personal protective equipment for its health care workers. In such an environment, transmissions from COVID-infected patients to health care workers to patients who are in hospitals for other reasons could play a major role in disease transmission. Suppose that this transmission resembled that in a ten-group uniform matching model with $R_0 = (6, 1, \ldots, 1)$. The most-active group in this model could represent the health care workers. In the early stages of the epidemic any non-health care worker who is infected will infect on average one other, 0.4 health care workers and 0.6 non-health care workers. An infected health care worker will in turn infect six others, again with 40%-60% split between health care workers and others. If a homogeneous SIR model is fit to early growth of such an epidemic one would estimate $\hat{R}_0 = \overline{R}_0 = (6^2 + 1^2 + \ldots + 1^2)/(6 + 1 + \ldots + 1) = 45/15 = 3$ and infer that herd immunity will not be reached until two-thirds of the population is infected. In fact, herd immunity can be reached much more easily. The key is to stop the within-hospital transmission. If five-sixths of the health care workers are immune, then each new infection

[15] Acemoglu et al. (2020) also include a discussion of targeting which point in the herd immunity region the system reaches.

Covid Economics 53, 23 October 2020: 1-32

will lead to just one other. The health care workers are just one-tenth of the population, so we can reach herd immunity with just 8.3% of the population having been infected.

In the model with homophilic matching we achieve herd immunity by choosing $S_1, \ldots, S_N$ such that $hS_iR_{0i} < 1$ for all i and so that

$$\sum_i \frac{R_{0i}}{\sum_k R_{0k}} \frac{S_iR_{0i} - 1}{1 - hS_iR_{0i}} \leq 0.$$

Here, the herd-immunity point with the lowest total number infected again involves having a lower fraction susceptible in the more active groups, but the solution will typically not be to fully concentrate the infected. Although the initial change in the constraint from reducing S_i away from one is largest in the most active group, the marginal benefit of reducing the fraction susceptible decreases as the fraction susceptible in a group is reduced, which may make the solution interior in multiple populations.

Achieving herd immunity with homophilic matching is more difficult than achieving herd immunity with uniform matching. This follows directly from the contrapositive of the result noted at the end of section 2.3: if a disease free state is stable for any h', then it must also be stable for all $h < h'$. This implies that the herd immunity region for a model with homophily parameter h', i.e. the set of S^0 for which $(S^0, 0)$ is stable, is a subset of the herd immunity region for a model with parameter h. The minimum fraction of the population that must have been infected to achieve herd immunity is therefore monotonically increasing in h. Finding the minimum threshold is very easy in the $h = 1$ case: the model is essentially a set of separate homogeneous SIR models so the solution is simply to set $S_i = \text{Min}(1, 1/R_{0i})$ in each subpopulation. In our five-population example with $R_0 = (3.5, 1.5, 1, 0.5, 0.5)$ this involves infecting 2/7 of those in subpopulation 1, 1/3 of those in subpopulation 2, and no others, which is the 21% of the total population mentioned earlier. For intermediate h one needs to solve the maximization problem described above, but we know the threshold increases continuously from 14% to 21% as h goes from 0 to 1. For $h = 0.5$ it is 15.5%.

An important factor to keep in mind when thinking about implications of results on herd immunity is that heterogeneous SIR models, like homogeneous SIR models, always "overshoot" their herd immunity thresholds in an uncontrolled epidemic. In the homogeneous SIR model, overshooting occurs because many are infected when herd immunity is reached and the infection is then reproducing at an approximately contant rate. For example, a homogenous SIR model with $R_0 = 16/7 \approx 2.3$ reaches its herd immunity threshold when just $1 - 7/16 \approx 56\%$ of the population has been infected but the infection will eventually hit about 87% of the population in an uncontrolled epidemic. In heterogeneous

Covid Economics 53, 23 October 2020: 1-32

Covid Economics 53, 23 October 2020: 1-32

SIR models, uncontrolled spread infects more people than are infected in the minimal herd immunity state for two reasons: the same overshooting effect as before, and because the path of the infection does not concentrate infections in the high-activity population to as great a degree as does a path that enters the herd immunity region at the minimal-infection point. This additional source of excess infections can be extremely potent. For example, whereas the five-population uniform-matching SIR model with $R_0 = (3.5, 1.5, 1, 0.5, 0.5)$ (which has $\overline{R}_0 = 16/7$) can be in the herd-immunity region with as little as 14% of the population infected, an infection starting from a small evenly distributed mass of infected will not reach the herd immunity region until 33% of the population is infected, and overshooting will result in 54% eventually being infected.

While I noted above that the minimal-infection herd immunity point entails more infections when matching is more homophilic, overshooting can be less extreme in homophilic models and when epidemics spread in an uncontrolled manner this can more than offset the difference in the herd immunity thresholds. For example, with the same R_0 vector as above, the fraction eventually infected is 46% with $h = 0.5$, 42% with $h = 0.75$, and just 31% with $h = 1$.

4.2 Overestimation of the difficulty of controlling an epidemic

While heterogeneous population models suggest that reaching herd immunity need not involve nearly as many infections as homogeneous SIR models suggest, they also suggest that avoiding herd immunity via selective lockdown policies may not be as difficult as homogeneous SIR models suggest.

Several recent papers have discussed optimal policies using frameworks in which transmission rates constant at time t can be reduced to $R_0(1 - x_t)$ by "locking down" a fraction x_t of the population.[16] This can reduce the fraction infected before a vaccine is developed, and reduce excess deaths from exceeding hospital capacity. In a homogeneous SIR model, lockdown policies that keep the population from reaching herd immunity incur large economic costs because the fraction infected will grow unless we keep the initial x_0 large enough so that $1 - x_0$ is below the herd immunity threshold. As a result, some optimal-policy simulations suggest that we may mostly want to use lockdowns just to temporarily slow the epidemic when hospitals would otherwise be overwhelmed.

The lower herd immunity thresholds of homogeneous models imply that targeted permanent lockdowns could keep the fraction infected from ever expanding by locking down a smaller fraction of the population. For example, in the $R_0 = (3, 1.5, 1.0, 0.5, 0.5)$ example discussed in the herd immunity

[16] See Acemoglu et al. (2020), Alvarez, Argente, and Lippi (2020), and Rowthorn and Toxvaerd (2012).

Covid Economics 53, 23 October 2020: 1-32

section, the problem of determining the minimum fraction the population that must be permanently locked down to keep the epidemic from ever expanding is mathematically equivalent to earlier calculation of the minimal herd-immunity threshold. Hence, the epidemic can be stopped by permanently locking down 14% of the population in the uniform matching case or at most 21% of the population in the homophilic model.

Temporary lockdowns can also be appealing in heterogeneous population models because they can serve as a means to guide the system toward a more desirable part of the herd-immunity region and/or reduce overshooting For example, to prevent the dramatic overshooting noted in the previous section, one could lock down all members of the lowest-activity populations once prevalence there reached a fraction of a percent, keep them locked down as the infection spreads through the most active populations, and then release them from lockdown once the population is close to herd immunity.

Figure 3 provides a numerical illustration. The solid blue series is the time path of new daily cases in a homogeneous population with $R_0 = 16/7 \approx 2.3$. The dashed red line is the time path of new daily cases in a heterogeneous uniform matching population with $R_0 = (3.5, 1.5, 1.0, 0.5, 0.5)$. Recall that the heterogeneous population has $\hat{R}_0 = 16/7$ and indeed the two series initially look identical. The infection in the heterogeneous population reaches herd immunity sooner and many fewer people are eventually infected than in the homogeneous population. But the lower damage absent a lockdown does not mean that the incremental benefit from a temporary lockdown is lower. The gray dashed line shows the path of the infection under a temporary targeted lockdown: we reduce activity by 20% in the highest activity populations and by 60% in the lower activity populations for a 60 day period. This reduces overshooting in the highest-activity population and reduces the number of low-activity people who are infected by members of the high-activity population as it is going through its peak. In the numeric example, it reduces the fraction who are ever infected from 54% to 38%.[17]

4.3 Overestimation of the impact of social distancing policies and endogenous behavioral responses

Cumulative US COVID-19 deaths grew roughly exponentially throughout March, passing 5000 on April 1st. Growth subsequently slowed dramatically. Many have noted that both government-mandated policies and endogenous individual reactions would be expected to contribute to the

[17] In the homogeneous model, implementing the same policy would make cases decline during the lockdown period, but there would be a massive second wave after the policy is lifted and the eventual total infected would only be reduced by about ten percentage points.

Covid Economics 53, 23 October 2020: 1-32

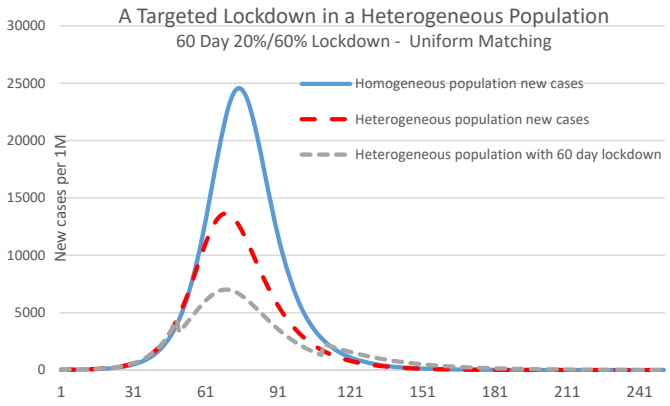


Figure 3: Effect of a temporary targeted shutdown in a heterogeneous population. The figure graphs new daily cases in three models. The dashed red line is a heterogeneous SIR model uniform matching and $R_0 = (3.5, 1.5, 1.0, 0.5, 0.5)$. The dashed gray line graphs cases for the same population assuming a temporary 60-day lockdown is imposed during the peak. The solid blue line is a homogenous SIR model with $R_0 = 16/7$ with no lockdown.

Covid Economics 53, 23 October 2020: 1-32

change.[18] Understanding the causal impact of each factor is critical to forecasting the impact of reopenings. A third effect also contributes to slowing growth in SIR models—the growth rate $\gamma(R_0 S(t) - 1)$ decreases as $S(t)$ declines—but this effect will be small in homogeneous SIR models calibrated to current conditions because $S(t)$ remains close to one.[19] In heterogeneous SIR models, however, this third effect can be nontrivial even in the early stages of an epidemic, particularly if matching is homophilic.

The effect is easiest to quantify in the uniform matching model. If the susceptible fraction has been reduced to S and the infection is still small, the growth of the infection will resemble that of a homogeneous SIR model with parameter $\overline{R}_0(S) = \sum_i \frac{R_{0i}}{\sum_k R_{0k}} S_i R_{0i}$. Writing $\overline{S}$ for the average fraction susceptible, the dominant eigenvector implies that the relative frequencies in the early infected population will be roughly proportional to their activity levels so $S_i \approx 1 - N\frac{R_{0i}}{\sum_k R_{0k}}(1-\overline{S})$. Differentiating with respect to $\overline{S}$ we find $\frac{d\overline{R}_0(\overline{S})}{d\overline{S}} = N\sum_i w_i^2 R_{0i}$, where we have written $w_i \equiv \frac{R_{0i}}{\sum_k R_{0k}}$ for the fraction of early infections which are in population i. Focusing just on the effect due to reductions in the most active group we have

$$\frac{d\log\overline{R}_0(\overline{S})}{d\overline{S}} = \frac{N\sum_i w_i^2 R_{0i}}{\sum_i w_i R_{0i}} \geq \frac{N w_1^2 R_{01}}{\sum_i w_i R_{0i}} = w_1 \cdot \frac{w_1 R_{01}/\sum_i w_i R_{0i}}{1/N}.$$

Note that the first term in the product is population 1's share of early infections and the second is the ratio of population 1's contribution to $\overline{R}_0$ to its share of the total population. In extreme examples where almost all early infections are in one small subpopulation, this effect can be very large. For example, $w_1 \approx 1$ in a model in which population 1 is just $1/N$ of the total population we have $\frac{d\log\overline{R}_0(\overline{S})}{d\overline{S}} \approx N$, i.e. the apparent $\overline{R}$ will have been reduced by about N% by the time 1% of the population has been infected. (In a homogeneous SIR model, the reduction in the apparent R_0 would be 1%.) The effect is smaller in uniform-matching models with less extreme heterogeneity in R_{0i}. For example, in the $R_0 = (3.5, 1.5, 1.0, 0.5, 0.5)$ example I have used frequently, $w_1 = \frac{1}{2}$, $\frac{w_1 R_{01}}{\sum_i w_i R_{0i}} \approx \frac{3}{4}$, and $1/N = 0.2$, so $\frac{d\log\overline{R}_0(\overline{S})}{d\overline{S}} \approx 2$. This suggests that the apparent $\overline{R}_0$ will have been reduced by about 10% when cumulative infections have reached 5%. This is larger than the 5% prediction of a homogenous SIR model, but is not a dramatic difference.

The reduction in the apparent $\overline{R}_0$ can be much larger in models with homophilic matching because

[18] See Baqaee et al. (2020), Farboodi, Jarosch, and Shimer (2020), Fernández-Villaverde and Jones (2020), Jones, Philippon, and Venkateswaran (2020), and Kudlyak, Smith, and Wilson (2020). Epidemiological estimates of changes in growth rates include Miller et al. (2020) and Unwin et al. (2020).

[19] A recent study in Sweden indicated that despite their embrace of herd immunity the fraction with antibodies is just 7% in Stockholm.

Covid Economics 53, 23 October 2020: 1-32

infections and loss of susceptibility are both more concentrated in the highest activity groups. For a simple illustration, think of a model with $h \approx 1$. Here, the power of exponential growth means that if we start from a tiny fraction infected in each group we will soon have almost all of the infected in the highest-activity group. As a result, we can perceive the growth process early in the epidemic to be close to R_{01} growth. If the infection peaks and declines in population 1 before it reaches a substantial size in population 2, the apparent growth rate can temporarily fall to well below one even though the epidemic is still in its early stages. Growth will then rise back to look like R_{02} growth in a second wave, and so on. Such nonmonotonic growth rates only occur when h is very close to one, but the fairly rapid early decline in the apparent growth rate as the epidemic burns itself out in the highest-activity population is a feature that persists well away from the $h = 1$ limit. If we take $h = 0.7$ in the $R_0 = (3.5, 1.5, 1.0, 0.5, 0.5)$ example, growth that looks like $\overline{R}_0 \approx 3$ growth early in the epidemic will slow to what looks like $\overline{R}_0 \approx 2.5$ growth by the time 5% of the population has been infected. Almost 20% of the highest-activity population is no longer susceptible at this point, and this substantially reduces the epidemic growth rate.

The slowdown of an epidemic continues as it approaches and passes the herd immunity threshold. Hence, viewing an epidemic in light of a homogeneous SIR model can both lead one to to mistakenly conclude that initial behavior changes were more effective than they were at slowing the epidemic and that later reopenings caused less acceleration than they did.

4.4 Underestimation of heterogeneity in R_0 across regions

The SIR parameter R_0 reflects both the contagiousness of a disease and the frequency and closeness of interactions in a population. It seems natural that R_0 should be larger in some countries or states than in others. For example, we might expect it to be larger in more densely populated and highly urbanized Belgium than in Sweden. But few economic analyses incorporate heterogeneity in R_0 across regions. This presumably reflects at least in part that the early epidemiological literature did not provide clear evidence of cross-country or cross-region differences. For example, Flaxman et al. (2020) provided estimates for 11 European countries from the period before lockdowns went into effect, and the 50% credible interval for Sweden (roughly 3.7–4.3) overlaps with the 50% credible intervals for 9 of the other 10 countries including Belgium.[20]

[20]Unwin et al. (2020) estimates a more flexible model with more recent data and reports much more substantial heterogeneity across US states, as do Fernández-Villaverde and Jones (2020). There is also a substantial range in early estimates of the rate at which COVID-19 spread in China.

Covid Economics 53, 23 October 2020: 1-32

The limited heterogeneity in reported R_0 parameters could reflect what is estimated when one applies a homogeneous SIR model to a heterogeneous world with homphilic matching. As an illustration, suppose that the differences between two countries lie not in differences between activity vectors $(R_{01}, R_{02}, \ldots, R_{0N})$, but in the fact that country a has a higher fraction of its population in the high activity groups than does country b. For example, it may be that both countries have working class subpopulations living in crowded urban housing and riding public transportation to jobs where they work in close proximity to others and rural populations with much lower contact rates, with the primary cross-country difference being in the relative fractions in each group. In an extreme homophilic model with $h \approx 1$, an estimation of R_0 would yield identical estimates of $\hat{R}_0 = R_{01}$ in both countries, regardless of whether important differences in the population compositions were present.

Again, these differences persist well away from the $h = 1$ limit. For example, with $h = 0.5$ a model with five equal sized populations with $R_0 = (3.5, 1.5, 1.0, 0.5, 0.5)$ will resemble $\overline{R}_0 \approx 2.75$ growth early in the epidemic, whereas a model with he same R_0 vector, but in which the three most active populations are each 10% of the population rather than 20% will resemble $\hat{R}_0 = 2.5$ growth. Homogeneous SIR epidemics with $R_0 = 2.75$ and $R_0 = 2.5$ follow similar paths – herd immunity is reached when 64% are infected in one model vs. 60% in the other and with overshooting the epidemics eventually infect 93% and 90%. It would be natural to not bother to incorporate such differences in an economic analysis of homogeneous SIR-based models. But the two heterogeneous models follow quite different paths with one eventually infecting 49% of the population and the other eventually infecting 29%. Accounting for the potential impacts of such differences seems much more important.

4.5 Misestimation of when epidemics start

A number of early papers fitting SIR models produced estimates of when epidemics started. In addition to satisfying intellectual curiosity, one motivation for such an exercise is that it may provide evidence on the size of the asympotomatic population. Features of the heterogeneous SIR model suggest that it will be very difficult to produce reliable estimates via such a method. Specifically, while I emphasized earlier that a heterogeneous SIR model will appear to grow at a rapid pace $\overline{R}_0$ from quite early on, this is not true at the very, very beginning. The infection only starts to grow at a rate related to the largest eigenvalue once the pattern of the distribution of infections across the populations is aligned with the principal eigenvector. Before this occurs, the growth rate can be very different depending on whether the initial infections are in a low- or high-activity population. This makes early growth rates unpredictable, and makes inferences about when an epidemic started very imprecise.

Covid Economics 53, 23 October 2020: 1-32

One of the most influential and inaccurate early papers on the COVID-19 epidemic may have misled in part this reason. Lourenço et al. (2020) calibrated an SIR model to estimate the fraction of the UK and Italy populations who were already infected as of March 19. In the three primary scenarios included in Figure 1, they estimate that the start of the UK epidemic occurred about 30 days before the first reported death, and then project forward to estimate that between 36% and 67% of the UK population was already infected as of March 19. One reason for the inaccuracy is that the analysis assumed that the death rate was much lower than now appears to be the case. Another source of the inaccuracy, however, may be another pair of assumptions—that deaths do not occur until well after infection and that the time series of infections followed an SIR path with R_0 equal to 2.25 or 2.75 from the very beginning.

In addition to being imprecise, homogeneous SIR-based inferences about epidemic origins may be biased. Growth rates were probably lower in the very early days than they were by the time the epidemic grew to the size where estimates of R_0 were first made. This difference may help reconcile why the fraction that antibody tests indicate have ever been infected is not larger, despite revelations that there was a case in France in late December and a death in California on February 6.[21] It may also help account for why some models, e.g. that shown in Figure 3 of Baqaee et al. (2020), find it difficult to match data on deaths from very early in the epidemic.[22]

5 Implications and Conclusions

The most basic message of this paper is that thinking about an epidemic in terms of homogeneous SIR models can lead to mistaken conclusions if the interactions are better described by a model with heterogeneous contact rates. Incorporating at least some heterogeneity need not be so difficult—in many cases what is being done with a single population model could be done quite similarly in a multipopulation model. But the remarkable pace at which the economics literature on COVID-19 has been progressing makes keeping up with the state of the art sufficiently difficult that my primary hope is that others will take the "heterogeneity matters" message to heart and incorporate it in their work.

Early in any epidemic, there is a great deal of scientific uncertainty about the disease transmission process. This paper's most important message about the COVID-19 epidemic itself is that to the

[21] Worobey et al. (2020) provide genome-based evidence that later early cases were not part of the main epidemics in Washington and Italy.

[22] Data inaccuracies may, of course, also be relevant here, so it is possible that the model predictions are closer to the truth than are the data.

Covid Economics 53, 23 October 2020: 1-32

extent that the epidemiological literature is suggesting that heterogeneity in transmission may be important, economists should be cognizant that we may still not understand its dynamics very well. Estimates of R_0 derived from observations in the early days of the epidemic may not reflect how the epidemic would spread now absent restrictive policies, and it is particularly difficult to estimate the parameters describing how the virus is spreading in less active communities. These parameters are critical to understanding how the epidemic may progress as restrictions are loosened. As questions about the impact of reopening policies become most salient, recognizing our limitations and doing our best to estimate the hard-to-estimate parameters is important. The greater speed with which the apparent R_0 can decline in heterogeneous models, particularly when matching is homophilic, also suggests that there may be more uncertainty than has been assumed in estimates of the impact both of distancing policies and of reopenings. The natural directions of bias are that we may overstate the impact that initial shutdown policies had in slowing the spread of COVID-19 and underestimate the extent to which the partial relaxations have accelerated the spread. It is particularly important to keep these biases in mind when estimates obtained in some region are used to provide advice to others.

A more optimistic implication of heterogeneous SIR models is that the COVID-19 epidemic may not be as bad as some models suggest. Models using growth rates estimated in the early days of the epidemic may overstate how rapidly the epidemic would have spread absent government intervention even if people had not taken it upon themselves to socially distance. And it is possible that epidemic growth can be slowed by herd immunity effects at prevalence levels substantially lower than naïve models suggest. If so, the option of reaching herd immunity, becomes less unattractive, particularly if the herd immunity level being contemplated is that which applies when cost-effective mitigation measures, such as universal mask wearing, are maintained, and if extensive efforts are made to keep infections out of vulnerable populations along the path. The possibility that the impact of restrictive policies may have been overestimated also suggests that some partial reopenings may be less damaging than anticipated.

Another important conclusion, however, is that the optimistic message that reaching herd immunity may not be as damaging as feared should not be taken to imply that trying to reach herd immunity is more advisable than earlier analyses suggest. Models with heterogeneity also suggest that controlling the spread of COVID-19 may be easier than thought. For one thing, benefits similar to those which herd immunity provides can be obtained by implementing targeted measures to prevent high-contact people, e.g. health care and nursing home workers, those riding public transportation, etc., from ever

being infected. This makes measures such as ensuring nursing home workers have adequate personal protective equipment even more powerful and cost-effective. And heterogeneous models also suggest that both permanent and temporary targeted lockdowns may be more effective as a means to limit the spread of the epidemic than homogeneous SIR models suggest. The good news on both sides of the equation makes it possible that correctly accounting for heterogeneity in the epidemic process could bolster the case for keeping in place policies that shut down high spread activities. Obviously, it would be valuable to know more about the nature of contact heterogeneity (and about the long-run health consequences of COVID-19 for survivors) to make this assessment.

I also noted that estimating SIR models on data early in an epidemic may lead one to underestimate the extent to which critical parameters of the epidemic process differ across regions. The changes in the course of epidemics in the aftermath of severe lockdown policies indicate that in aggregate policies and behavioral changes had a very large impact on R_0.[23] The resurgences we have seen as restrictions have been loosened and individuals become less vigilant make clear that we are still quite far from herd immunity and we will need to retain some restrictions for quite some time, perhaps well beyond when vaccines become available. The limits to what is safe, however, could be very different in different locations and at different times. It would be valuable to have tailored guidance so that we do not simply have to rely on trying to infer the effect of each set of incremental changes.

The recent economics literature includes several papers examining multipopulation SIR models including Acemoglu et al. (2020), Baqaee et al. (2020), and Favero, Ichino, and Rustichini (2020). While the analyses in these papers have not been calibrated to fully capture within age-group heterogeneity in contact rates, they certainly could move in this direction. Baqaee et al. (2020), for example, could in theory have "simply" used age $\times$ occupation groups instead of age groups as the basis of their model, replacing 5×5 matrices with 330×330 matrices, to capture contact heterogeneity across those working in each of the sectors they consider. This still, however, would not have captured within-occupation heterogeneity.

In addition to the computational challenges, a factor that will limit our ability to calibrate more complex models is the limited data that is available on heterogeneity in contact rates. Just as serology data has helped compensate for the weak identification of asymptomatic cases in regular SIR models, more data may also provide the solution to the weak identification problem noted here. Although several firms have already made location tracking data available to researchers, privacy concerns have

[23] See Chernozhukov, Kasaha, and Schrimpf (2020) and Goolsbee and Syverson (2020) for estimates of the importance of policies and self-motivated behavioral changes.

Covid Economics 53, 23 October 2020: 1-32

limited public releases to means within various cells. One simple step that could potentially greatly enhance the value of this data is to also release within-cell variances and within-individual time series correlations. While those developing apps that use Bluetooth interactions to track phone-to-phone proximity are rightly being careful with privacy, they could potentially provide an even more valuable source of information on contact distributions. Epidemiologists are also able to exploit variation in virus genomes to provide more micro-based estimates of disease-transmission.[24] While economists are unlikely to have the expertise to take advantage of genomic data, keeping current on insights coming out of these analyses will be important.

Covid Economics 53, 23 October 2020: 1-32

[24]See, for example, Miller et al. (2020) and Worobey et al. (2020).

Covid Economics 53, 23 October 2020: 1-32

References

[1] Daron Acemoglu, Victor Chernozhukov, Iván Werning, and Michael D Whinston. "Optimal Targeted Lockdowns in a Multi-Group SIR Model". NBER Working Paper No. 27102. 2020.

[2] Fernando Alvarez, David Argente, and Francesco Lippi. "A Simple Planning Problem for COVID-19 Lockdown". *Covid Economics* 14 (2020), pp. 1–32.

[3] Viggo Andreasen and Freddy B Christiansen. "Persistence of an infectious disease in a subdivided population". *Mathematical Biosciences* 96.2 (1989), pp. 239–253.

[4] Andrew Atkeson. "How deadly is COVID-19? Understanding the difficulties with estimation of its fatality rate". NBER Working Paper No. 26965. 2020.

[5] Christopher Avery, William Bossert, Adam Clark, Glenn Ellison, and Sara Fisher Ellison. "Policy implications of models of the spread of coronavirus: Perspectives and opportunities for economists". *Covid Economics* 12 (2020), pp. 21–68.

[6] David Baqaee, Emmanuel Farhi, Michael J Mina, and James H Stock. "Reopening Scenarios". NBER Working Paper No. 27244. 2020.

[7] Tom Britton. "Estimation in multitype epidemics". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.4 (1998), pp. 663–679.

[8] Tom Britton, Frank Ball, and Pieter Trapman. "A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2". *Science* (2020).

[9] David Champredon, Michael Li, Benjamin M Bolker, and Jonathan Dushoff. "Two approaches to forecast Ebola synthetic epidemics". *Epidemics* 22 (2018), pp. 36–42.

[10] Victor Chernozhukov, Hiroyuki Kasaha, and Paul Schrimpf. "Causal Impact of Masks, Policies, Behavior on Early Covid-19 Pandemic in the US". *arXiv preprint arXiv:2005.14168* (2020).

[11] Nikolaos Demiris and Philip D O'Neill. "Bayesian inference for stochastic multitype epidemics in structured populations via random graphs". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.5 (2005), pp. 731–745.

[12] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. "On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations". *Journal of Mathematical Biology* 28.4 (1990), pp. 365–382.

[13] Jonathan Dushoff and Simon Levin. "The effects of population heterogeneity on disease invasion". *Mathematical Biosciences* 128.1-2 (1995), pp. 25–40.

[14] Martin S Eichenbaum, Sergio Rebelo, and Mathias Trabandt. "The macroeconomics of epidemics". NBER Working Paper No. 26882. 2020.

[15] Maryam Farboodi, Gregor Jarosch, and Robert Shimer. "Internal and external effects of social distancing in a pandemic". *Covid Economics* 9 (2020), pp. 22–58.

[16] Carlo A Favero, Andrea Ichino, and Aldo Rustichini. "Restarting the economy while saving lives under Covid-19". CEPR Discussion Paper No. DP14664. 2020.

[17] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. 2020". *DOI* 10 (2020), p. 77482.

[18] Jesús Fernández-Villaverde and Charles I Jones. "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities". NBER Working Paper No. 27128. 2020.

[19] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Unwin, Helen Coupland, T Mellan, Harisson Zhu, T Berah, J Eaton, P Perez Guzman, et al. "Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries" (2020).

[20] M Gabriela M Gomes, Ricardo Aguas, Rodrigo M Corder, Jessica G King, Kate E Langwig, Caetano Souto-Maior, Jorge Carneiro, Marcelo U Ferreira, and Carlos Penha-Goncalves. "Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold". *medRxiv* (2020).

[21] Austan Goolsbee and Chad Syverson. "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020". *Becker-Friedman Working Paper 2020-80* (2020).

[22] Michael Greenstone and Vishan Nigam. "Does social distancing matter?" *Covid Economics* 7 (2020), pp. 1–22.

[23] Herbert W Hethcote. "The mathematics of infectious diseases". *SIAM Review* 42.4 (2000), pp. 599–653.

[24] Matthew O. Jackson and Dunia Lopez-Pintado. "Diffusion and contagion in networks with heterogeneous agents and homophily". *Network Science* 1.1 (2013), pp. 49–67.

[25] John A Jacquez, Carl P Simon, and James S Koopman. "Core Groups and the R0s for Subgroups in Heterogeneous SIS and SI Models". *Epidemic Models: Their Structure and Relation to Data* 5 (1995), p. 279.

[26] Callum J Jones, Thomas Philippon, and Venky Venkateswaran. "Optimal mitigation policies in a pandemic: Social distancing and working from home". NBER Working Paper No. 26984. 2020.

[27] William Ogilvy Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics". *Proceedings of the Royal Society of London. Series A* 115.772 (1927), pp. 700–721.

[28] Petra Klepac, Adam J Kucharski, Andrew JK Conlan, Stephen Kissler, Maria L Tang, Hannah Fry, and Julia R Gog. "Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project" (2020).

[29] Ivan Korolev. "Identification and Estimation of the SEIRD Epidemic Model for COVID-19". Binghamton University Working Paper No. 3569367. 2020.

[30] Mariana Kudlyak, Lones Smith, and Andrea Wilson. "For whom the bell tolls: avoidance behavior at breakout in an SI3R model of covid". *Virtual Macro Seminar*. 2020.

Covid Economics 53, 23 October 2020: 1-32

[31] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. "Superspreading and the effect of individual variation on disease emergence". *Nature* 438 (2005), pp. 355–359.

[32] José Lourenço, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig Thompson, Peter Simmonds, Paul Klenerman, and Sunetra Gupta. "Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic". *medRxiv* (2020).

[33] Robert M May and Roy M Anderson. "The transmission dynamics of human immunodeficiency virus (HIV)". *Applied Mathematical Ecology*. Springer, 1989, pp. 263–311.

[34] Danielle Miller, Michael A Martin, Noam Harel, Talia Kustin, Omer Tirosh, Moran Meir, Nadav Sorek, Shiraz Gefen-Halevi, Sharon Amit, Olesya Vorontsov, et al. "Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel". *medRxiv* (2020).

[35] Jo el Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases". *PLOS Medicine* 5.3 (2020), pp. 381–391.

[36] Adriano A Rampini. "Sequential lifting of covid-19 interventions with population heterogeneity". NBER Working Paper No. 27063. 2020.

[37] Bob RE Rowthorn and Flavio Toxvaerd. "The optimal control of infectious diseases via prevention and treatment". CEPR Discussion Paper No. DP8925. 2012.

[38] James H Stock. "Coronavirus Data Gaps and the Policy Response to the Novel Coronavirus". *Covid Economics* 3 (2020), pp. 1–11.

[39] H Juliette T Unwin, Swapnil Mishra, Valerie C Bradley, Axel Gandy, Michaela Vollmer, Thomas Mellan, Helen Coupland, Kylie Ainslie, Charles Whittaker, Jonathan Ish-Horowicz, et al. "State-level tracking of COVID-19 in the United States" (2020).

[40] Pauline Van den Driessche and James Watmough. "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission". *Mathematical Biosciences* 180.1-2 (2002), pp. 29–48.

[41] Cécile Viboud, Kaiyuan Sun, Robert Gaffey, Marco Ajelli, Laura Fumanelli, Stefano Merler, Qian Zhang, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, et al. "The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt". *Epidemics* 22 (2018), pp. 13–21.

[42] Michael Worobey, Jonathan Pekar, Brendan B Larsen, Martha I Nelson, Verity Hill, Jeffrey B Joy, Andrew Rambaut, Marc A Suchard, Joel O Wertheim, and Philippe Lemey. "The emergence of SARS-CoV-2 in Europe and the US". *bioRxiv* (2020).

Covid Economics 53, 23 October 2020: 1-32

Is deforestation spreading COVID-19 to the indigenous peoples?[1]

Humberto Laudares[2]

Date submitted: 16 October 2020; Date accepted: 20 October 2020

This paper examines deforestation's effect on the COVID-19 transmission to indigenous peoples and its transmission mechanisms. To that end, I analyze the Brazilian case and use new datasets that cover all the country's municipalities daily. Relying on a fixed-effects model, I find that deforestation is a powerful and consistent variable to explain the transmission of COVID-19 to indigenous populations. The estimates show that one unit increase in deforestation per 100 km² is associated, on average, with the confirmation of 2.4 to 5.5 new daily cases of COVID-19 in indigenous people 14 days after the deforestation warnings. One km² deforested today results in 9.5% more new COVID-19 cases in two weeks. In accumulated terms, deforestation explains at least 22% of all COVID-19 cases confirmed in indigenous people until 31 August 2020. The evidence suggests that the main mechanisms through which deforestation intensifies human contact between indigenous and infected people are illegal mining and conflicts.

Covid Economics 53, 23 October 2020: 33-71

1 Pedro Henrique Gagliardi provided outstanding research assistance. I had valued suggestions from Felipe Valencia and from the anonymous reviewer. All remaining errors are mine.

2 Ph.D. in Development Economics, Graduate Institute of International and Development Studies, Geneva, and Ph.D. candidate in Global Health, University of Geneva.

Covid Economics 53, 23 October 2020: 33-71

1 Introduction

The COVID-19 pandemic has grown exponentially in the Developing World. Until the end of August 2020, Brazil ranked second in the number of COVID-19 cases and the death toll, lagging behind the United States. Brazil turned in 2020 not only a worldwide epicenter of COVID-19 but also of deforestation. While the negative externalities of deforestation are well documented in the literature, less is known about how deforestation can affect the transmission of COVID-19 to vulnerable ethnic groups, such as the indigenous peoples[1], enlarging existing income and racial inequality gaps.

Brazil had more than 3.8 million confirmed COVID-19 cases and 120 thousand deaths by the end of August 2020[2]. It represents 15.1% of the confirmed cases and 14.3% of the total deaths reported globally[3]. At the same time, deforestation has increased by 25% from January-June 2020 (3.070 km^2) in comparison to the same period in 2019[4]. 55% of the deforested lands this year have been also burned (Moutinho et al. (2020)). On top of it, deforestation had dramatically expanded in indigenous lands, the *de facto* forest's main guardian (Laudares (2016), Baragwanath and Bayi (2020)), while coronavirus infected more than 20 thousand and victimized more than 800 indigenous people.

This paper asks whether deforestation has been a key driver in the COVID-19 transmission to indigenous peoples. It also focuses on exploring the channels through which deforestation may affect the spread of the disease to this ethnic group relative to the others.

To that end, I construct a daily panel with 5,417 municipalities from 1 March 2020 to 31 August 2020 with COVID-19 cases and deforestation data. I use a fixed-effects model to exploit the effects of deforestation on COVID-19's morbidity of indigenous populations. The independent variables are lagged in 5 or 14 days, following the clinical evidence of asymptomatic period after the contamination. This method is particularly interesting to analyze such big panel data because it captures municipality within variation and controls for the effects of time-invariant variables. I also conduct the empirical analysis in a cross-section format because I can add more covariates

[1]The WHO Executive-Director recently highlighted that 'indigenous peoples often have a high burden of poverty, unemployment, malnutrition and both communicable and non-communicable diseases, making them more vulnerable to COVID-19 and its severe outcomes.' – United Nations news, July 2020

[2]Retrieved at 23 September 2020 from COVID-19 Dashboard - Johns Hopkins University

[3]Retrieved at 21 September 2020 from WHO Coronavirus Disease (COVID-19) Dashboard

[4]Data from the Real Time Deforestation Detection System (Deter) of the National Institute for Space Research - INPE

Covid Economics 53, 23 October 2020: 33-71

and explore the potential mechanisms through which deforestation affects COVID-19 transmission in indigenous peoples.

Deforestation is a powerful and consistent variable to explain the transmission of COVID-19 to indigenous populations. Relying on the fixed-effects model, I find that one unit increase in deforestation per 100 Km^2 is associated, on average, with the confirmation of 2.1 to 2.4 new daily cases of COVID-19 in indigenous people 14 days after the deforestation warnings. If I add nonlinearity in the model, the coefficient jumps to 5.5. The deforestation per Km^2 that takes place in $t = 1$ will increase the COVID-19 cases among indigenous people by 9.5% fourteen days later ($t = 15$). Using weekly panel data, a unit increase in deforestation warnings per 100 Km^2 elevates the new COVID-19 by 30% two weeks after the event.

My main cross-section results, based on a state-fixed effects estimation, show that one unit change in warning areas for deforestation per 100 km^2 within the Amazon Forest and the Cerrado ecosystem in Brazil increases the number of COVID-19 cases confirmed in originary peoples by 55. A straightforward linear calculation suggests that, on average, deforestation explains at least 22% of all COVID-19 cases confirmed in indigenous people until 31 August 2020. Population density and economic inequality are the most relevant control variables correlated with coronavirus transmission among originary peoples.

Under the 'bad controls' framework (Angrist and Pischke (2008)), I also test the key transmission mechanisms – namely, wildfires, cattle ranching, illegal mining, and conflicts involving indigenous people – as controls. The evidence suggests that the two strongest mechanisms through which deforestation affects the spread of COVID-19 in indigenous communities are illegal mining and conflicts. But deforestation explains a large part through which illegal mining (84 to 91%) and conflicts (81 to 97%) contribute to new COVID-19 cases of indigenous people, considered that, as a 'bad control', it affect the spread of COVID-19 through other channels as well.

I run the robustness checks regression deforestation in COVID-19 hospitalizations, following the primary panel data's same econometric approach. The effect of deforestation on COVID-19 hospitalizations is not as direct as is the case of COVID-19 transmission. Besides, the clinical development of the patient requires to take into consideration additional individual characteristics. In light of those circumstances, I use the hospitalization data as a proxy for COVID-19 incidence. While this dataset from the Ministry of Health provides the possibility to compare COVID-19 hospitalizations by race, there are much fewer observations for indigenous peoples. Using this database, I find that deforestation is only positively correlated – and statistically sig-

Covid Economics 53, 23 October 2020: 33-71

nificant – with COVID-19 hospitalization in indigenous people, but not with other races. I estimate that 9.14% of all COVID-19 hospitalizations of indigenous people relate to deforestation.

A larger literature explores the negative externalities of deforestation on the economy and society (Nordhaus (2019), Malhi et al. (2008), Castro et al. (2019)). A growing amount of evidence shows policy (Souza-Rodrigues (2019) Assunção et al. (2019)) Burgess, Costa, and Olken (2019) Chimeli and Soares (2017)) , political (Pailler (2018)), and economic forces (Sonter et al. (2017)) play a role in deforestation in Brazil, especially in the Amazon region. A growing stream of works has been developing on the impact of the COVID-19 on racial (Bertocchi and Dimico (2020), McLaren (2020)), gender (Alon et al. (2020)), opportunity (Bacher-Hicks, Goodman, and Mulhern (2020)) and economic inequalities (Campello, Kankanhalli, and Muthukrishnan (2020)). In the case of Brazil, Baqui et al. (2020) and Bruce et al. (2020) evaluated the impact of COVID-19 in different ethnicities using the Ministry of Health database[5] that does not incorporate the indigenous health statistics, since they are not harmonized.

This paper innovates in analyzing an additional negative externality of deforestation in the context of a pandemic affecting indigenous populations. As a result, it also shed light on a specific mechanism on how ethnic and health inequalities have been deepening as the pandemic develops, and the government fails to respond.

To my knowledge, this is the first paper that estimates the relationship between deforestation and the spread of COVID-19 affecting indigenous peoples. Also, this is the first paper that uses the COVID-19 datasets published by the Special Department of Indigenous Health (SESAI)[6]) at the Ministry of Health and the Articulation of Indigenous Peoples of Brazil (APIB), the Brazilian indigenous peoples' major representative organization.

I believe the findings are of relevance to policymakers as well. Ending deforestation – and fighting all the illegal economic activities related to it – brings enormous benefit for the climate and the environment as a whole and contributes to curb the transmission of COVID-19 among indigenous populations.

The rest of the paper is organized as follows. The next section presents the background of the context in which indigenous peoples in Brazil are dealing with increasing deforestation and COVID-19 transmission. I then detail the data and the empirical strategy in Sections 3 and 4. Subsequently, I will present the empirical results, followed by sections 6 and 7 on the transmission mechanisms and robustness check. I

[5]SIVEP-Gripe (Sistema de Informação de Vigilância Epidemiológica da Gripe)
[6]Secretaria Especial de Saúde Indígena

conclude with the main takeaways of the study.

2 Indigenous peoples, deforestation, and COVID-19

There are 311 indigenous peoples living in Brazil, totaling 760 thousand people (0.36% of the total population). As of 31 August 2020, the COVID-19 pandemic affected 158 of those communities. The COVID-19 cases in the indigenous people represent 0.6 to 0.8% of the total[7]. In the Amazon region, the indigenous mortality is the highest among all the ethnic groups (Fellows et al. (2020)), while for the whole country, 'pardos' and black people present the highest coronavirus death toll (Baqui et al. (2020)).

However, the existing comparison of COVID-19 transmission – not mortality – among races using Brazil as a case study underestimates the effects of COVID-19 on the indigenous peoples (Baqui et al. (2020), Bruce et al. (2020)). The main reason is that indigenous public health department statistics do not integrate the universal health system's statistics[8]. The Ministry of Health assistance to indigenous communities is run separately from the universal health system and decentralized in the Indigenous Special Sanitary Districts (ISSD[9]). There are 34 ISSDs in Brazil, and their borders do not follow the country's original administrative boundaries. The government reports COVID-19 statistics of the indigenous peoples exclusively at the ISSD level, which aggregates 219 municipalities and several indigenous peoples.

Besides, the access of indigenous communities to health equipment is scarce. Given the level of severity observed in the case of COVID-19[10] and the geographical barriers, isolated indigenous communities face relevant obstacles to reach on time specialized health facilities or intensive care units. On top of it, there are 120 communities uncontacted, and 76% of those have not been confirmed yet. The existing concern is the chance of illegal miners, missionaries, or illegal 'land grabbers' transmit COVID-19 to the uncontacted indigenous peoples[11].

All the factors mentioned above may increase the sub-notification of COVID-19 cases and delay their reporting. Oviedo et al. (2020) and Azevedo et al. (2020) estimate

[7]The lower bound is the Ministry of Health data and the upper bound, the The Articulation of Indigenous Peoples of Brazil (APIB) data.

[8]Sistema Unico de Saúde, SUS, as it is called in Portuguese.

[9]The acronym in Portuguese is DSEIs

[10]Reuters/G1, 15 May 2020

[11]Data from Instituto Socioambiental (ISA)

Covid Economics 53, 23 October 2020: 33-71

the vulnerability of the indigenous communities to COVID-19, where they have also considered demographic and infrastructure aspects. The Articulation of Indigenous Peoples of Brazil (APIB)[12], a representative organization of the indigenous peoples, highlights other reasons of sub-notification, such as racism, misreporting, and lack of transparency of the official authorities (APIB (2020)).

The Amazon region detains the highest concentration of COVID-19 cases, hospitalization, and indigenous deaths for COVID-19. The states of Amazonas and Pará represents 38% of the Covid-19 cases reported in indigenous people. Map 1 also displays that this is the region where the highest deforestation incidence[13].

While shreds of evidence show that deforestation might be one of the key variables to explain the spread of COVID-19 among indigenous communities, analysts, and indigenous representation bodies point out other variables as well. They are: illegal mining[14], land grabbing and timber lodgers[15], cattle ranching and meat processing plants[16], and transport through the rivers[17]. Moreover, health workers[18] and missionaries[19] pose a potential risk of the spread of COVID-19 among indigenous people considering the imminent contact with their communities.

However, the association between deforestation and transmission of COVID-19 is not automatic. COVID-19 is a disease transmitted primarily by droplets from coughing, sneezing, or even talking (WHO (2020)). Therefore, although human proximity is required to transfer the disease, there are still several reasons to believe that deforestation is related to the pathogen's spread.

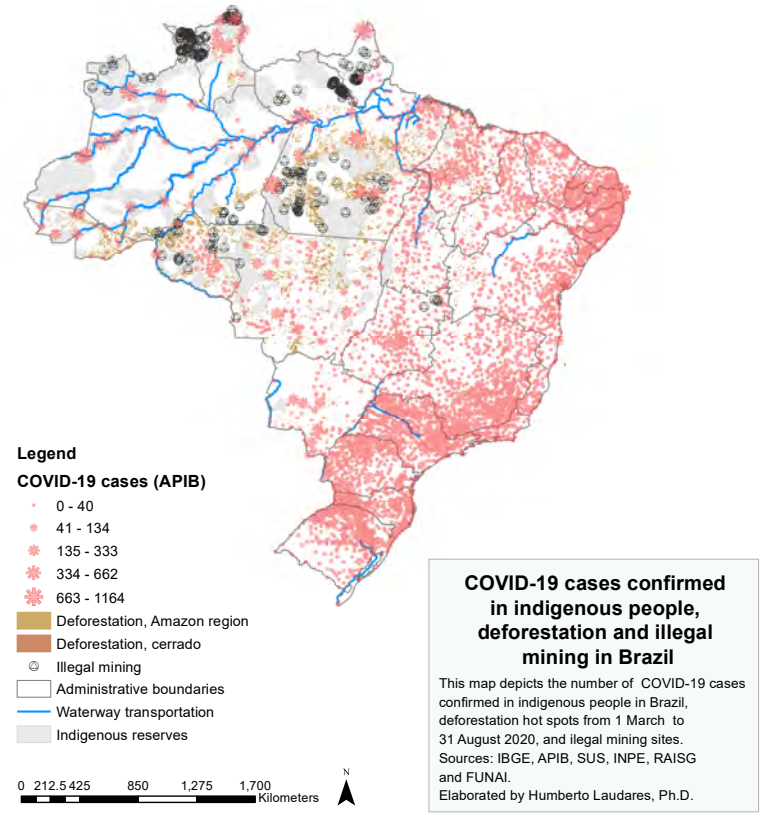
First, 72% of the deforested lands in 2020 are in conservation areas and indigenous lands[20], which entails some level of – peaceful or violent – social interaction. APIB (2020) reports compelling cases of how deforestation can disentangle in conflicts. Second, whether deforestation targets land grabbing, cattle ranching, illegal mining, or timber extraction, indigenous communities are already exposed to the virus through improper contact with infected people[21]. Socioambiental (2020), for instance, argues the threat imposed by deforestation for the Yanomami people is highly dangerous due

[12] Articulação dos Povos Indígenas do Brasil (APIB)
[13] According data from the Deter system, Terrabrasilis, INPE
[14] Data from Amazonia Socioambiental
[15] G1, 4 August 2020
[16] Globo Rural, 10 June 2020
[17] BBC, 8 May 2020
[18] Instituto Socioambiental, 24 July 2020
[19] The Economist, 9 July 2020
[20] Folha de S. Paulo, Mining and Deforestation, 25 June 2020
[21] O Globo, Ianomami, mining and COVID-19, 2 June 2020

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

Figure 1: COVID-19 cases confirmed in indigenous people, deforestation and illegal mining in Brazil



to a greater vulnerability to COVID-19 exposure. The conflicts and close interactions between miners and indigenous have transmitted COVID-19 to about 40% of Yanomami people. Third, deforestation puts pressure on indigenous people to have a forced displacement[22] to regions where the virus may already be present. Finally, Oliveira et al. (2020) and Rocha and Sant'Anna (2020) argue that the fires from the increasing deforestation, combined with the drought and wildfires, worsens respiratory health risks, including the COVID-19 cases, increasing the demand for health services and the locomotion to cities.

Based on the evidence mentioned above, the following section briefly describes the data used in the empirical analysis. It also explains the identification strategy used to show the effects of deforestation on the spread of COVID-19 in indigenous communities.

Covid Economics 53, 23 October 2020: 33-71

3 Data

In this section, I describe the different sources and levels of aggregation of the main variables used in the empirical analysis. The Annex details all other variables used as controls.

3.1 Panel data

The panel has 184 days and 5,417 municipalities, totaling 969,728 observations. It starts on 1 March, when the pandemic officially started in Brazil, and ends on 31 August 2020. The two data sources of COVID-19 cases confirmed in indigenous people used in the paper are from the Special Department of Indigenous Health (SESAI)[23] at the Ministry of Health, and the Articulation of Indigenous Peoples of Brazil (APIB).

SESAI is responsible for collecting data from indigenous people, and the reporting is apart from the Brazilian universal health system. The datasets are at Indigenous Special Sanitary Districts (ISSD) level, which are decentralized administrative healthcare units dedicated to the indigenous peoples. There are 34 ISSDs within the borders of 219 municipalities. The APIB's estimates are based on SESAI's statistics, but the organization adds more information collected through the indigenous networks and local governments countrywide and subtracts the duplicated data. APIB's objective,

[22] National Congress, 27 June 2020

[23] In Portuguese, Secretaria Especial de Saúde Indígena

Covid Economics 53, 23 October 2020: 33-71

in this case, is to reduce the sub-notification of COVID-19 cases in indigenous people (APIB (2020)).

The APIB argues lack of transparency on SESAI statistics – essential to deal with the pandemic – and testing for indigenous peoples. The SESAI's dataset, for instance, does not inform whether the indigenous people that contracted COVID-19 lives in indigenous lands, non-urban settings, or not. It also does not notify which indigenous peoples the diagnosed people belong to. The absence of this information hampers how these communities plan in dealing with the pandemic themselves. It makes the task to match the data of SESAI very challenging with other Ministry of Health's datasets. However, SESAI and APIB databases are the best available information to study the spread of COVID-19 in indigenous populations.

To this end, as a first step, I convert the data from ISSD to municipal level, based on the size of the indigenous population living in each municipality relative to the ISSD total. The Annex section details the data conversion to the municipal level.

For robustness, I also use the COVID-19 hospitalizations of indigenous people from the Ministry of Health's *SIVEP-Gripe* database as a proxy for COVID-19 incidence, and reports the daily data from the Brazilian universal health system (SUS[24]). However, there are fewer indigenous peoples' observations than the other two panels because the *SIVEP-Gripe* does not include data from SESAI. I believe that most of the data is from indigenous living in the cities, even though there seems to be an overlap between the datasets[25]. Besides, the *SIVEP-Gripe* dataset reports COVID-19 hospitalizations by race. The relationship between deforestation and COVID-19 hospitalization tends to be weaker than the cases reported. However, the main advantage of working with this data as proxy is the possibility to compare the results with other ethnic groups, namely black, white, 'pardo' (mixed), and East Asian (yellow) people.

On the right side of the equation, deforestation, the main independent variable, is measured by the warning areas of deforestation in $100km^2$ within the Amazon Forest and the Cerrado (Brazilian Savannah) ecosystems. The data, collected from Brazil's National Institute for Spatial Research, is at the municipal level from 1 March to 31 August 2020.

Table 1 exhibits the summary statistics of the daily data, totaling 996,876 observations.

[24]Sistema Único de Saúde

[25]In the case of deaths reported, APIB (2020) estimates an overlap around 41%, while 10.8% of the total is not clear if it is an overlap or not.

Table 1: Summary statistics: daily panel data

Variables	N	Mean	SD	Min	Max	Sum
COVID-19 confirmed cases of indigenous people (SESAI)	996,876	0.023	0.519	0	90	23,179
COVID-19 confirmed cases of indigenous people (APIB)	996,876	0.0291	.822	0	251	28,985
COVID-19 hospitalizations of indigenous people	996,728	0.001	0.042	0	8	1,163
COVID-19 hospitalizations of black people	996,728	0.016	0.308	0	40	15,527
COVID-19 hospitalizations of white people	996,728	0.109	1.667	0	248	108,942
COVID-19 hospitalizations of 'pardo' people	996,728	0.112	1.378	0	151	111,290
COVID-19 hospitalizations of East Asian people	996,728	0.004	0.096	0	14	3,754
Deforestation (per $100km^2$)	996,728	0.000	0.004	0	0.7	116.2
Log Deforestation (per km^2)	7,053	-0.729	1.453	-8.517	4.324	

Notes to Table 1. Table 1 displays the descriptive statistics of the key variables used in the panel data estimations. The COVID-19 confirmed cases of indigenous peoples' data are published by the Ministry of Health's Special Department of Indigenous Health (SESAI)[26]) and by the Articulation of Indigenous Peoples of Brazil (APIB), the Brazilian indigenous peoples' major representative organization. The COVID-19 hospitalization data from the Ministry of Health's *SIVEP-Gripe* database is used in the robustness check. Deforestation data is extracted from the Real-Time Deforestation Detection System (Deter) of the National Institute for Space Research - INPE.

I also collapse the daily data to weekly to smooth the noise. In the case of COVID-19 cases in indigenous people is even more relevant due to potential delays in the reporting because of geographical distances and low access to health equipment. Table 7 displays the summary statistics of the weekly data.

3.2 Cross-section

The main dependent variable used in the cross-section analysis is the accumulated numbers of COVID-19 cases reported by the Special Department of Indigenous Health (SESAI) at the Ministry of Health and the total cases compiled by APIB. The deforestation data, the primary independent variable, is presented in accumulated values from 1 March to 31 August 2020.

The benefit of using a cross-section at the municipal level is adding several other control variables and capturing the mechanisms through which deforestation affects the spread of COVID-19 in indigenous communities. Besides, adding relevant controls also minimizes the potential bias derived from omitted variables.

Covid Economics 53, 23 October 2020: 33-71

CEPR PRESS

Covid Economics 53, 23 October 2020: 33-71

Table 2: Summary statistics: cross-section

Variables	N	Mean	SD	Min	Max
COVID-19 cases of indigenous people (SESAI)	5,417	4.286	36.723	0	1,025
COVID-19 cases of indigenous people (APIB)	5,417	5.350	47.176	0	1,164
COVID-19 hospitalizations of indigenous people	5,417	0.215	2.021912	0	64
COVID-19 hospitalizations of black people	5,417	2.867	39.884	0	2,367
COVID-19 hospitalizations of white people	5,417	20.115	255.001	0	17,044
COVID-19 hospitalizations of 'pardo' people	5,417	20.548	181.798	0	9,134
COVID-19 hospitalizations of East Asian people	5,417	0.693	10.889	0	738
Deforestation (per $100km^2$)	5,417	0.020	0.172	0	5.552
Population density	5,385	123.405	637.898	0.049	14,208
Illegal mining	5,385	0.008	0.0920	0	1
Conflict involving indigenous people (CPT)	5,417	0.037	0.188	0	1
Conflict involving indigenous people (CIMI)	5,417	0.040	0.197	0	1
Cattle ranching	5,334	31,664	72,910	21	1571,245
Wildfires - Fire Radiative Power (FRP)	5,417	1,179	12,493	0	506,161
GDP	5,385	9.749	0.678	8.097	12.750
Inequality (Gini coefficient)	5,380	0.503	0.066	0.284	0.808
Extreme poverty	5,384	0.279	0.448	0	1
Number of emergency rooms	5,384	0.059	0.366	0	13
Access to public roads	3,503	3.983	1.795	-4.605	12.496
Proximity to waterway	5,385	0.055	0.227	0	1
Proximity to an environmental protection agency (Ibama)	5,385	0.937	0.597	0	4.012
Access to clean water	2,558	0.677	0.240	0	1
Access to treated sewage	2,558	0.350	0.470	0	1
Rainfall	5,040	11.634	5.335	0	33
Distance to the coast	5,084	0.048	0.213	0	1
Distance to the state capital	5,299	251.472	164.103	0	1,476
Altitude	5,299	411.488	293.885	0	1,628
Latitude	5,385	-16.49	8.30	-33.68	4.60
Longitude	5,385	-46.27	6.445	-72.89	-34.81
Types of soil:					
Sglei	5038	0.009	0.096	0	1
Slat	5038	0.338	0.473	0	1
Sluvi	5038	0.038	0.192	0	1
Sneo	5038	0.130	0.336	0	1
Snit	5038	0.030	0.171	0	1
Splan	5038	0.040	0.196	0	1
Splint	5038	0.036	0.188	0	1

Notes to Table 2. Table 2 exhibits all variables included in the cross-section analysis. The Annex brings detailed information about each of them.

The control variables include access to clean water, public roads and waterways[27], population density, GDP, income inequality, and geographical variables, such as rainfall, altitude, and distance to the state capital (see in the Annex for a detailed description of the variables). I also use illegal mining, conflicts involving indigenous people, wildfires, and cattle ranching as control variables, even if they are also the main mechanisms connecting deforestation with the dissemination of COVID-19 in indigenous communities.

Table 2 exhibits the summary statistics of the variables used in the cross-section estimates.

4 Empirical Strategy

The empirical analysis exploits the relationship between deforestation and COVID-19 cases from 1 March 2020 to 31 August 2020[28]. The following econometric model will be the main reference in analyzing a municipal level and daily panel data. I estimate equations of the form:

$$COVID_{it} = \alpha + \rho COVID_{i(t-l)} + \beta \Gamma_{i(t-l)} + \delta_t + \lambda_i + \upsilon_{it}, \quad (1)$$

where $COVID_{it}$ is the dependent variable that captures the number of COVID-19 cases of indigenous people in the municipality i in period t. The lagged variable of COVID-19 cases in l time units on the right-hand side is included to reflect the disease transmission mechanism's intrinsic persistence. $\Gamma_{i(t-l)}$ is the main explanatory variable, namely the lagged value of deforestation alerts per 100 Km^2. The parameter β measures the causal effect of deforestation on the transmission of COVID-19 in indigenous people. Additionally, λ_i is the set of municipality dummies and δ_t the time effects related to common trends in deforestation. The error term is expressed by $\upsilon_{i(t)}$, absorbing all other omitted effects. Time is expressed by t, and l is the lagged values.

The fixed-effects model measures the municipality within variation over time. The parameter λ_i captures time-invariant municipality unobserved characteristics that affect deforestation, avoiding the potential problem of omitted variables. This is the key difference between the pooled ordinary least squares (OLS) model and the fixed-effects.

I use the lagged values (of 5 and 14 days) of the main independent variables in the estimations. There are two reasons for that. First, the empirical literature on public

[27]Waterways are the main transportation modal in the Amazon region.

[28]The first cases of COVID-19 in indigenous communities started to be reported from 1 April. However, since I use lagged variables for the independent variable, the data starts from 1 March 2020

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

health reports that individuals infected with coronavirus may remain asymptomatic between 5 to 12 days (Lauer et al. (2020)). The authors' findings show that 97.5% of patients develop symptoms within 11.5 days of infection. On the other hand, the WHO states that the incubation period of the coronavirus is, on average, 5 to 6 days, but it can be up to 14 days[29]. Second, as mentioned before, the geographic barriers and difficult access to hospitals by indigenous people delay the reporting of contaminated indigenous people's health status. Also, SESAI reports the COVID-19 cases by the day of the diagnosis, not the contamination day. In the Amazon region, where there is the highest concentration of indigenous peoples, the main transportation system is fluvial. A trip to a medium city can take more than a day. Besides, using lagged variables eliminates any potential contemporaneous effects between Covid-19 reported cases and deforestation.

I also estimate the relationship mentioned above using a cross-section, by accumulating both COVID-19 cases and deforestation values for the whole period analyzed and adding other relevant covariates. The main estimations based on the cross-section data rely on the state-fixed effects model, that captures within-state variation, λ_i, while time dimension δ_t remains constant because $t = 1$. The specification is particularly interesting because the states – together with the municipalities – are the key federal entities that implement policies to combat the COVID-19 in Brazil. There was significant heterogeneity in the policies adopted across the country. Therefore, estimating the state-fixed effect and the other covariates added tend to be more efficient in reducing the omitted variables bias than the OLS.

The cross-section data allow us to investigate the mechanisms that reinforce the relationship between deforestation and the spread of COVID-19 to originary peoples, such as illegal mining, cattle ranching, wildfires, and conflicts involving indigenous people, under the 'bad control' framework (Angrist and Pischke (2008) and Cinelli, Forney, and Pearl (2020)). In this case, the mentioned mechanisms can also be the outcome of deforestation. Controlling for them would introduce bias in the estimation results and lead to the 'bad control' problem. However, 'bad controls' are useful to provide insights about the mechanisms through which the independent variable affects the dependent one, once compared the estimations that contemplate and do not include them (Maccini and Yang (2009).

The next section exhibits the main results of my analysis.

[29]See at WHO COVID-19 Situation Report 73

5 Results

5.1 Panel data results

Figure 1 exhibits a visual format of the main results. Table 3 reports the main results of the panel data using municipal level fixed-effects estimation. Throughout the paper, the standard errors used are robust against heteroskedasticity and clustered at the municipal level to prevent serial correlation. Columns 1 to 4 have as dependent variable COVID-19 cases reported by the Ministry of Health (SESAI), and columns 5 to 8 exhibit COVID-19 cases reported by the Indigenous Peoples association (APIB). The independent variable, deforestation, is lagged in 14 days. Table 8 reports the same estimation using 5 days lagged independent variables. The first and the fifth columns report pooled OLS results, while the others exhibit fixed effects estimations.

Table 3 implies that one unit increase in deforestation per 100 Km^2 is associated, on average, with the confirmation of 2.1 to 2.4 new daily cases of COVID-19 in indigenous people 14 days after the deforestation warnings. As a reference, the OLS coefficient is 3.9 for the cases reported by SESAI (column 1) and 4.7 when using the cases consolidated by APIB (column 5), both with 14 days lag deforestation as an independent variable. Adding nonlinearity in the model, columns (3) and (7) report one unit increase in deforestation areas per 100 Km^2 explaining 5.1 and 5.5 new COVID-19 cases in indigenous people respectively[30]. Table 3 also shows that deforestation, which takes place in $t = 1$, will increase the COVID-19 cases among indigenous people by 3.5 (column 4) to 9.5% (8) fourteen days later ($t = 15$).

The reason to add nonlinearity in columns (3) and (7) is natural and geographical barriers, such as rivers and mountains, and infrastructure. I would expect that the sign of the coefficient of deforestation is positive, while the squared value is negative, as in table 3 holds.

I find that lagged values of COVID-19 cases reported in indigenous people within a given municipality, since the disease is highly contagious (Petersen et al. (2020)), explain 14 to 30% of the new cases of COVID-19 reported in the same ethnic group. The implicit cumulative effect of deforestation on COVID-19 dissemination[31] coefficients is negative and lower than one.

Performing the estimations using weekly data, I lose variability, however, it smooths

[30]For reference, the median is 0.004. The maximum points for deforestation are that up to 0.28 (SESAI) and 0.33 per 100 Km^2 (APIB)

[31]The implicit cumulative effect of deforestation coefficient is estimated by Deforestation per 100 Km^2_{t-l}/(1-COVID-19 cases$_{t-l}$). This variable was based on Acemoglu et al. (2008)

Covid Economics 53, 23 October 2020: 33-71

the data and accounts for potential delays in reporting COVID-19 cases in indigenous peoples. I use lagged independent variables in one and two weeks. Table 9 reports one unit increase in deforestation per 100 Km^2 is associated, on average, with the confirmation of 6.3 to 10.3 new weekly cases of COVID-19 in indigenous people one or two weeks after the deforestation warnings (columns 1, 2, 5 and 6). The quadratic models in columns (3) and (7) show that one unit increase in deforestation areas per 100 Km^2 explains the COVID-19 transmission to 25 (SESAI) to 33 (APIB) indigenous people per week. Alternatively, columns (4) and (8) suggest that one unit increase in deforestation warnings per 100 Km^2 increases the weekly transmission of COVID-19 among indigenous peoples by 15.4% (SESAI) to 30% (APIB).

Table 3: Fixed-effects results: deforestation and COVID-19 cases in indigenous peoples

	COVID-19 cases SESAI				COVID-19 cases APIB			
	Pooled OLS	Fixed effects	Fixed effects	Fixed effects	Pooled OLS	Fixed effects	Fixed effects	Fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deforestation (per 100 $Km^2)_{t-14}$	3.912***	2.400**	5.056***		4.818***	2.076***	5.476***	
	(1.258)	(0.945)	(1.943)		(1.292)	(0.733)	(1.684)	
Log Deforestation (per $Km^2)_{t-14}$				0.0350*				0.0948***
				(0.0197)				(0.0269)
Deforestation (per 100 $Km^2)^2_{t-14}$			-7.657**				-9.804***	
			(3.147)				(3.469)	
COVID-19 cases $SESAI_{t-14}$	0.415***	0.296***	0.296***	0.265***				
	(0.0561)	(0.0522)	(0.0522)	(0.0569)				
COVID-19 cases $APIB_{t-14}$					0.227***	0.140***	0.139***	0.0986
					(0.0330)	(0.0331)	(0.0331)	(0.0731)
Implied cumulative	-0.894	-0.885*	-0.828*	-0.793*	-0.915**	-0.773*	-0.698	-0.794*
effect of deforestation	(0.478)	(0.481)	(0.471)	(0.467)	(0.466)	(0.456)	(0.456)	(0.413)
Constant	0.0159***	-0.00003	-0.00004	0.223**	0.0199***	-0.00007	-0.00008	0.451***
	(0.00210)	(0.00216)	(0.00216)	(0.0938)	(0.00229)	(0.00327)	(0.00327)	(0.139)
Observations	920,874	920,876	920,876	6,495	920,876	920,839	920,839	6,494
R-squared	0.151	0.083	0.083	0.100	0.051	0.021	0.021	0.067
Number of municipalities		5,417	5,417	913		5,417	5,417	913

Notes to Table 3. Columns 1 and 5 present pooled OLS estimations with robust standard errors clustered by municipality in parentheses. The remaining columns are fixed-effects estimation at municipal level with time and municipality dummies and robust standard errors clustered at municipal level in parentheses. The implicit cumulative effect of deforestation coefficient is estimated by Deforestation per 100 Km^2_{t-l}/(1-COVID-19 $cases_{t-l}$). All the independent variables are lagged in 14 days. The standard errors are in parentheses, where *** p<0.01, ** p<0.05, * p<0.1.

Covid Economics 53, 23 October 2020: 33-71

CEPR PRESS

Covid Economics 53, 23 October 2020: 33-71

5.2 Cross-section results

This section presents the results of cross-section data. Table 4 shows the estimations using OLS (columns 1 and 5) and state fixed effects (columns 2, 3, 4, 6, 7 and 8), allowing for within state variation. In columns 4 and 8, I add as controls the main mechanisms that correlate with both the COVID-19 cases and deforestation variables (see table 6). While they can be called 'bad controls', equations' coefficients might be biased when they are included, thus they are important as a reference in the analysis.

The columns (3) and (7) exhibit my main estimations because they include 'good' but not 'bad' controls, which reduces the potential bias found in both OLS equations (columns 1 and 5), and the state fixed-effects estimations without the incorporation of key covariates (columns 2 and 6). Thus, although I opted to include fewer control variables, there are quite relevant controls for the analysis[32].

The main results imply that one unit change in warning areas for deforestation per 100 km^2 within the Amazon Forest and the Cerrado ecosystem in Brazil increases COVID-19 cases by 37.39 (SESAI) to 55.22 (APIB) indigenous people. As expected, the coefficients are smaller than the OLS baseline (columns 1 and 5) but quite similar to the state-fixed effects baseline (columns 2 and 6).

Until the end of August, the warning areas of deforestation totaled 11,622 Km^2. Doing a straightforward linear calculation means that, on average, 4,345 to 6,418 indigenous people could have contracted COVID-19 due to the deforestation, based on SESAI and APIB data, respectively. In other words, deforestation explains at least 18.7 to 22.1% of all COVID-19 cases confirmed in indigenous people until 31 August 2020.

Population density and economic inequality are key variables to explain the spread of COVID-19 in Developing Countries (Ahmed et al. (2020), Pequeno et al. (2020)), which is also the case here. Both variables are positively correlated with COVID-19 cases and statistically significant. One unit hike in the Gini coefficient, which captures income inequality, is associated with an increase of 67 to 86 news cases of COVID-19. The mentioned results are even more concerning in the context of a pandemic that will certainly enlarge income and opportunity gaps between the rich and the poor (Campello, Kankanhalli, and Muthukrishnan (2020) Blundell et al. (2020), Dorn, Cooney, and Sabin (2020), Vahidy et al. (2020)).

There are few mechanisms through which deforestation could enhance human contact and contribute to the spreading of the coronavirus, such as wildfires, cattle ranch-

[32]In the Annex, I present additional regressions with more controls for reference.

ing, illegal mining, and conflicts. All those variables correlate with both COVID-19 cases and deforestation. Not only they can serve as a transmission mechanism of deforestation, but these mechanisms can also contribute to more COVID-19 contamination independently of deforestation. For instance, illegal mining can be expanded through deforestation in indigenous lands (Sonter et al. (2017)), generating some sort of human contact between indigenous and non-indigenous. Also, COVID-19 can be spread through well established illegal mining activities independently of deforestation.

Columns (4) and (8) include the mentioned variables as 'bad controls'. While I will take a closer look at them in the next section, Table 4 pinpoints their correlation with COVID-19, once controlled for deforestation and remaining covariates. Only illegal mining and conflicts present the expected – and statistically significant at the 99% confidence interval – results. I find that the presence of illegal mining in a given municipality results in 122 to 160 cases of COVID-19.

Since there are 46 municipalities with reported illicit mining activities, I can deduce that, on average, illegal mining explains 22 to 25% of the COVID-19 in indigenous people in Brazil. Similarly, the existence of conflicts involving indigenous people, including land disputes with illegal miners and timber lodgers, in a given municipality is linked with 41 to 53 COVID-19 cases. Doing a similar calculation with the 199 registered, they could explain about 36% of the indigenous people's COVID-19 cases. While the illegal mining and conflict dummies are useful for the analysis, they present limitations. First, as a binary variable, they capture the average effect of the existence of those activities in a given municipality but not of their intensity as a continuous variable such as deforestation. Second, in 23 towns (50% of cities that posse illegal mining activities reported) the data overlaps, and the variables also correlate with each other.

However, even using the 'bad controls' as controls, the effects of deforestation on the transmission of COVID-19 to indigenous people is consistently positive and statistically significant. The magnitude of the coefficients drop by 55 to 61% but remains relatively large, explaining 8.5%, on average, of all COVID-19 cases that indigenous people contracted in Brazil. Nevertheless, an equation that included 'bad controls' is more useful to check how robust the baseline estimation is and understand its transmission mechanisms.

In the following section, I will discuss further the main mechanisms of transmission.

Covid Economics 53, 23 October 2020: 33-71

CEPR PRESS

Table 4: State-fixed effects (accumulated): deforestation and COVID-19 cases in indigenous peoples

	COVID-19 cases SESAI				COVID-19 cases APIB			
	OLS	State fixed effects	State fixed effects	State fixed effects	OLS	State fixed effects	State fixed effects	State fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deforestation ($100km^2$)	46.58***	37.20***	37.39***	16.67***	65.64***	55.09***	55.22***	21.46***
	(10.27)	(2.853)	(2.943)	(4.285)	(17.02)	(3.695)	(3.835)	(5.586)
Population density			0.00139*	0.00298***			0.00235**	0.00523***
			(0.000768)	(0.00113)			(0.00100)	(0.00147)
GDP			-0.479	-1.275			0.348	-0.713
			(1.016)	(0.963)			(1.324)	(1.255)
Inequality (Gini coefficient)			67.23***	48.05***			86.13***	62.09***
			(9.099)	(8.623)			(11.86)	(11.24)
Wildfires				-0.116**				-0.0257
				(0.0566)				(0.0737)
Cattle ranching				0.00290				-0.00174
				(0.00810)				(0.0106)
Illegal mining				121.7***				159.9***
				(5.961)				(7.772)
Conflict				41.63***				53.11***
				(2.720)				(3.546)
Constant	3.327***	3.519***	-35.18***	-18.55*	4.009***	4.225***	-55.37***	-33.89**
	(0.487)	(0.466)	(10.82)	(10.29)	(0.629)	(0.603)	(14.10)	(13.42)
Observations	5,417	5,417	5,040	4,992	5,417	5,417	5,040	4,992
R-squared	0.048	0.147	0.146	0.254	0.057	0.133	0.141	0.251
Geographical controls	No	No	Yes	Yes	No	No	Yes	Yes

Notes to Table 4. Columns 1 and 5 present OLS estimations with robust standard errors in parentheses. The remaining columns are fixed-effects estimation at state level and robust standard errors clustered at state level in parentheses. The geographical variables included are rainfall, distance to the coast, distance to the state capital, and altitude. The standard errors are in parentheses, where *** p<0.01, ** p<0.05, * p<0.1.

Covid Economics 53, 23 October 2020: 33-71

6 Mechanisms

This section focus on the transmission channels of the main effect. To this end, I use the cross-section data. Columns (4) and (8) of Tables 4 function as a reference for this section as well, since it estimates the effects of the set of main mechanisms – wildfires, illegal mining, cattle ranching, and conflicts – on COVID-19 cases reported in indigenous communities. I rely on the same type of estimations applied in the prior section: OLS and state-fixed effects.

Table 11 regresses the mechanisms variables on deforestation to estimate their joint

relationship with the main independent variable of the model. I consistently find a positive correlation between the mechanisms and deforestation. The coefficients of cattle ranching lose their statistical significance when other control variables are added to the model. Taking column (2) as a reference, I find that one percent change in wildfires and cattle ranching are associate with a 7.15% and 4.6% change in deforestation, and a municipality that posses illegal mining and conflict is associated, respectively, with 179.3% and 63.1%[33] change in deforestation.

On the other hand, table 12 exhibits the reversed calculation, estimating the effects of deforestation on the mechanisms. I confirm that the relationship between them is positive and statistically significant. Tables 13 and 14 exhibit the results of each mechanism regressed on COVID-19 cases without deforestation as an independent variable. I find positive and statistically significant values only for wildfires, illegal mining, and conflicts.

The evidence suggests that the two strongest mechanisms through which deforestation affects the spread of COVID-19 in indigenous communities are illegal mining and conflicts. In the subsections below, I comment on each of those two mechanisms separately, based on the 'bad controls' framework (Angrist and Pischke (2008)).

6.1 Illegal mining

Table 11 implies a high correlation between deforestation and illegal mining. Column (6) in table 14 shows that the existence of illegal mining in a given municipality is associated with 189 (column (6)) or 174.5 (column (9)) cases of COVID for indigenous people. The estimated coefficient for illegal mining in column (8) at table 4, which controls for deforestation, is 159.9.

I interpret this as an indication that deforestation contributes to the transmission of COVID-19 to originary peoples through illegal mining ($\approx$ 84 to 91%) and other potential mechanisms aside deforestation.

6.2 Conflicts

In the Brazilian setting, conflict is an intuitive mechanism through which deforestation can disseminate COVID-19 among indigenous communities. Simultaneously, it is intertwined with other mechanisms such as illegal mining, wildfires, or forced displacements[34]

[33] I followed the standard calculation: (exp(B) - 1)*100%

[34] I have not found data to test these mechanisms.

Covid Economics 53, 23 October 2020: 33-71

Table 11 shows that conflicts is correlated with deforestation. Column (8) in table 14 implies that the occurrence of conflicts involving indigenous peoples within the borders of a municipality is associated with 65.8 (column (8)) or 54.6 (column (9)) cases of COVID. Table 4, controlling for deforestation, estimates this parameter in 53.11.

My understanding is that deforestation explains a large part through which conflicts contribute to new COVID-19 cases of indigenous people (81 to 97%), but, as a bad control, I also recognize that it affects the spread of COVID-19 through other channels as well.

7 Robustness check

The available databases report COVID-19 cases by the 34 Indigenous Special Sanitary Districts (ISSD). According to Saúde Indígena) (2020), until 29 August 2020, 54.1% of the the notified COVID-19 cases were confirmed. 95% of the confirmed cases were based on laboratory tests[35], while the remaining cases were clinically diagnosed. About 378 (1.6%) of the cases resulted in death. SESAI's data does not report hospitalization rates.

Alternatively, the Ministry of Health's *SIVEP-Gripe* database reports COVID-19 hospitalizations by race. However, as mentioned before, there are very few indigenous peoples' observations compared to the other two panels used in this paper. Based on a shred of evidence, I believe that the Ministry of Health's *SIVEP-Gripe* dataset mostly contains information from indigenous living in the cities, although it is not certain how much both datasets overlap.

The effect of deforestation on COVID-19 hospitalizations is not as direct as is the case of COVID-19 transmission. Besides, the clinical development of the patient requires to take into consideration additional individual characteristics. In light of those circumstances, I use the hospitalization data as a proxy for COVID-19 incidence. Therefore, I rely on this data to check for robustness and compare the effect of deforestation on COVID-19 hospitalization among races, namely indigenous, black, white, 'pardo' (mixed), and East Asian (yellow) people.

Table 5 reports the main results of deforestation on hospitalizations by race, based on a daily and municipal level panel data, and on a fixed-effects model. The independent variables are lagged in 14 days, reproducing the same approach used in the table

[35]APIB (2020) argues that SESAI only uses the serological test (rapid tests), and not the gold standard COVID-19 real-time reverse transcription polymerase chain reaction (rRT-PCR) test.

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

3. I note that deforestation is only positively correlated – and statistically significant – with COVID-19 hospitalization in indigenous people. Column (2) implies that one unit increase in deforestation areas per 100 Km^2 at $t-14$ is associated with 0.05 COVID-19 hospitalizations of indigenous people.

Table 5: Fixed-effects results: deforestation and COVID-19 hospitalizations by race

	Indigenous		Black	White	'Pardo'	East Asian
	OLS	Fixed effects	Fixed effects	Fixed effects	Fixed effects	Fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is COVID-19 hospitalizations by race					
Deforestation (per 100 Km^2)$_{t-14}$	0.120***	0.0525***	0.0151	0.0200	0.184	-0.0102
	(0.0459)	(0.0112)	(0.0521)	(0.199)	(0.211)	(0.0206)
Indigenous$_{t-14}$	0.137***	0.0661***				
	(0.0160)	(0.0010)				
Black$_{t-14}$			0.487***			
			(0.0009)			
White$_{t-14}$				0.568***		
			(0.0008)			
'Pardo'$_{t-14}$					0.576***	
					(0.0008)	
East Asian$_{t-14}$						0.170***
						(0.0010)
	0.0011***	-0.0000	0.0008	0.0253**	0.0071	0.0011
	(0.00004)	(0.0005)	(0.0026)	(0.0100)	(0.0106)	(0.0010)
Observations	920,890	920,890	920,890	920,890	920,890	920,890
R-squared	0.019	0.005	0.237	0.322	0.337	0.030
Number of municipalities	5,417	5,417	5,417	5,417	5,417	5,417
5,417	5,417					

Notes to Table 3. Columns 1 exhibits a pooled OLS estimation with robust standard errors in parentheses. The remaining columns are fixed-effects estimation at the municipal level with time and municipality dummies and robust standard errors in parentheses. All the independent variables are lagged in 14 days. Columns 1 and 2 have as dependent variable COVID-19 hospitalizations of indigenous people, column 3, black, 4, white, 5, 'pardo' (mixed), and 6, East Asian ('yellow'). Following the Brazilian Institute of Geography and Statistics' classification and guidelines, the race self-declared by the patients. The standard errors are in parentheses, where *** p<0.01, ** p<0.05, * p<0.1.

Table 15 exhibits the accumulated data in a cross-section format. Columns (1) and (2) confirm the finding of table 5, where deforestation only presents a positive and statistically significant causal effect on COVID-19 of indigenous people, but not of the other races.

Until the end of August, the warning areas of deforestation totaled 11,622 Km^2. Based on the deforestation parameter estimated in column (2), I can infer that, on

average, 106 indigenous people are hospitalized because of COVID-19 due to deforestation. In other words, it represents 9.14% of all COVID-19 hospitalizations of indigenous people until 31 August 2020. The percentage is about half the magnitude found using SESAI and APIB's data.

At the same time, population density and inequality associate with COVID-19 hospitalizations across all ethnic groups. But the magnitude of the coefficients differs substantially for each race. Table 15 show that white people's coefficients are, one average, 422 times higher than for the indigenous people, 4.8 than for black people, and 0.15 than for 'pardos'. While richer municipalities tend to have, on average, a higher number of COVID-19 hospitalizations across all races, except for indigenous people, inequality of income is consistently correlated with COVID-19 hospitalizations.

Covid Economics 53, 23 October 2020: 33-71

8 Final considerations

This paper documents a positive and statistically significant relationship between deforestation and the transmission of COVID-19 in indigenous communities. This correlation, when using hospitalization as a proxy of COVID-19 incidence, was not found in other ethnic groups. Even in the context of the COVID-19 pandemic, deforestation and the intertwined expansion of illegal mining have been growing in Brazil – especially in the Amazon region and within the indigenous reserves – with the consent of the current central government. Consequential conflicts involving indigenous peoples, also within indigenous reserves, boost the transmission of COVID-19 in this vulnerable ethnic group.

Using new datasets, I find that deforestation explains about 22% 0f all COVID-19 cases confirmed in indigenous populations. One Km^2 deforested today results in 9.5% more new COVID-19 cases in two weeks.

I do not have the presumption to state that deforestation causes the growth of COVID-19 cases in native populations. Further work is necessary to reach that conclusion; however, I believe this paper brings a stepping-stone for additional work on this topic. Also, the fact that COVID-19 related statistics of indigenous peoples and data from private health system are not included in a common reporting system it is a problem in itself. The data harmonization on COVID-19 would be vital to track the disease's development on a more realistic base and better evaluate the effects of the pandemic on society.

I believe the presented results are policy-relevant. The evidence suggests that ending deforestation is an optimal environmental policy. Still, it is also a health and

economic key issue given the importance to curb the spread of the COVID-19 and decrease the intensity of the economic shocks the pandemic has been causing at micro and macro levels.

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

References

Acemoglu, Daron et al. (June 2008). "Income and Democracy". In: *American Economic Review* 98(3), pp. 808–42. DOI: 10.1257/aer.98.3.808. URL: https://www.aeaweb.org/articles?id=10.1257/aer.98.3.808.

Ahmed, Faheem et al. (May 2020). "Why inequality could spread COVID-19." In: *The Lancet. Public health* 5 (5), e240. ISSN: 2468-2667. DOI: 10.1016/S2468-2667(20)30085-2. ppublish.

Alon, Titan et al. (Apr. 2020). *The Impact of COVID-19 on Gender Equality*. COVID Economics, Working Paper 4. Center for Economic Policy Research. URL: https://cepr.org/sites/default/files/CovidEconomics4.pdf.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

APIB (2020). *Vidas Indígenas Importam! COVID-19 e Povos Indígenas, o enfrentamento das violências durante a pandemia*. Tech. rep. Articulação dos Povos Indígenas do Brasil (APIB).

Assunção, Juliano et al. (2019). *Optimal environmental targeting in the amazon rainforest*. Tech. rep. National Bureau of Economic Research.

Azevedo, M. et al. (2020). "Análise de Vulnerabilidade Demográfica e Infraestrutural das Terras Indígenas à Covid-19".

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern (July 2020). *Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time*. Working Paper 27555. National Bureau of Economic Research. DOI: 10.3386/w27555. URL: http://www.nber.org/papers/w27555.

Baqui, Pedro et al. (2020). "Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study". In: *The Lancet Global Health*.

Baragwanath, Kathryn and Ella Bayi (2020). "Collective property rights reduce deforestation in the Brazilian Amazon". In: *PNAS (University of Michigan, Ann Arbor,USA)* 117, pp. 20495–20502. ISSN: 0027-8424. DOI: 10.1073/pnas.1917874117.

Bertocchi, Graziela and Arcangelo Dimico (July 2020). *COVID-19, Race, and Redlining*. COVID Economics, Working Paper 38. Center for Economic Policy Research. URL: https://cepr.org/sites/default/files/CovidEconomics38.pdf#Paper4.

Blundell, Richard et al. (2020). "COVID-19 and Inequalities*". In: *Fiscal Studies* 41(2), pp. 291–319. DOI: 10.1111/1475-5890.12232. eprint: https://onlinelibrary.

wiley.com/doi/pdf/10.1111/1475-5890.12232. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-5890.12232.

Bruce, Raphael et al. (2020). "Racial Inequality in Health Care During a Pandemic". In: *Available at SSRN: https://ssrn.com/abstract=3691313*.

Burgess, Robin, Francisco Costa, and Benjamin A Olken (2019). "The Brazilian Amazon's Double Reversal of Fortune". In: *SocArXiv*.

Campello, Murillo, Gaurav Kankanhalli, and Pradeep Muthukrishnan (May 2020). *Corporate Hiring under COVID-19: Labor Market Concentration, Downskilling, and Income Inequality*. Working Paper 27208. National Bureau of Economic Research. DOI: 10.3386/w27208. URL: http://www.nber.org/papers/w27208.

Castro, Marcia C et al. (2019). "Development, environmental degradation, and disease spread in the Brazilian Amazon". In: *PLoS biology* 17(11), e3000526.

Chimeli, Ariaster B and Rodrigo R Soares (2017). "The use of violence in illegal markets: Evidence from mahogany trade in the Brazilian Amazon". In: *American Economic Journal: Applied Economics* 9(4), pp. 30–57.

Cinelli, Carlos, Andrew Forney, and Judea Pearl (2020). *A Crash Course in Good and Bad Controls*. Tech. rep.

Dorn, Aaron van, Rebecca E. Cooney, and Miriam L. Sabin (Apr. 2020). "COVID-19 exacerbating inequalities in the US." In: *Lancet (London, England)* 395 (10232), pp. 1243–1244. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(20)30893-X. ppublish.

Fellows, M. et al. (2020). *Não são números, são vidas! A ameaça da covid-19 aos povos indígenas da Amazônia brasileira*. Tech. rep. Instituto de Pesquisa Ambiental da Amazônia (IPAM) and Coordenação das Organizações Indígenas da Amazônia Brasileira (COIAB).

Laudares, Humberto (2016). *From Telegraphs to Space: TransportInfrastructure, Development and Deforestation inthe Amazon*. Tech. rep. The Graduate Institute of International and Development Studies.

Lauer, Stephen A. et al. (May 2020). "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application." In: *Annals of internal medicine* 172 (9), pp. 577–582. ISSN: 1539-3704. DOI: 10.7326/M20-0504. ppublish.

Maccini, Sharon and Dean Yang (June 2009). "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall". In: *American Economic Review* 99(3), pp. 1006–26. DOI: 10.1257/aer.99.3.1006. URL: https://www.aeaweb.org/articles?id=10.1257/aer.99.3.1006.

Covid Economics 53, 23 October 2020: 33-71

Malhi, Y. et al. (2008). "Climate Change, Deforestation, and the Fate of the Amazon". In: *Science* 319, pp. 169–172. ISSN: 0036-8075. DOI: 10.1126/science.1146961.

McLaren, John (June 2020). *Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data.* Working Paper 27407. National Bureau of Economic Research. DOI: 10.3386/w27407. URL: http://www.nber.org/papers/w27407.

Moutinho, P. et al. (2020). *The Amazon in flames: deforestations and fire during the COVID-19 pandemic". Tech. rep. Amazon Environmental Research Institute.*

Nordhaus, William (2019). "Climate Change: The Ultimate Challenge for Economics". In: *American Economic Review* 109, pp. 1991–2014. ISSN: 0002-8282. DOI: 10.1257/aer.109.6.1991.

Oliveira, Gabriel de et al. (2020). "Smoke pollution's impacts in Amazonia". In: *Science* 369(6504). Ed. by Jennifer Sills, pp. 634–635. ISSN: 0036-8075. DOI: 10.1126/science.abd5942. eprint: https://science.sciencemag.org/content/369/6504/634.2.full.pdf. URL: https://science.sciencemag.org/content/369/6504/634.2.

Oviedo, A.F. et al. (2020). "Povos indígenas e Covid-19: relato emergencial". In: ed. by R. Pacheco. Editora Primata. Chap. A vulnerabilidade socioambiental dos povos indígenas no Brasil ao Covid-19.

Pailler, Sharon (2018). "Re-election incentives and deforestation cycles in the Brazilian Amazon". In: *Journal of Environmental Economics and Management* 88, pp. 345–365. ISSN: 0095-0696. DOI: 10.1016/j.jeem.2018.01.008.

Pequeno, Pedro et al. (2020). "Air transportation, population density and temperature predict the spread of COVID-19 in Brazil". In: 8, e9322. ISSN: 2167-8359. DOI: 10.7717/peerj.9322.

Petersen, Eskild et al. (Sept. 2020). "Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics." In: *The Lancet. Infectious diseases* 20 (9), e238–e244. ISSN: 1474-4457. DOI: 10.1016/S1473-3099(20)30484-9. ppublish.

Rocha, Rudi and André Sant'Anna (2020). *Winds of Fire and Smoke: Air Pollution and Health in the Brazilian Amazon.* Tech. rep. Instituto de Estudos para Políticas de Saúde (IEPS).

Saúde Indígena), SESAI (Secretaria Especial de (2020). *Informe Epidemiológico 17: Doença por Coronavírus (COVID-19) em populações indígenas.* Tech. rep. Ministério da Saúde, Secretaria Especial de Saúde Indígena.

Socioambiental, Instituto (2020). *The impacts of the pandemic on the yanomami indigenous territory.* Tech. rep. Instituto Socioambiental.

Covid Economics 53, 23 October 2020: 33-71

Sonter, Laura J. et al. (2017). "Mining drives extensive deforestation in the Brazilian Amazon". In: *Science* 8. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00557-w.

Souza-Rodrigues, Eduardo (2019). "Deforestation in the Amazon: A unified framework for estimation and policy analysis". In: *The Review of Economic Studies* 86(6), pp. 2713–2744.

Vahidy, Farhaan S. et al. (Aug. 2020). "Racial and ethnic disparities in SARS-CoV-2 pandemic: analysis of a COVID-19 observational registry for a diverse US metropolitan population." In: *BMJ open* 10 (8), e039849. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2020-039849. epublish.

WHO (2020). *Transmission of SARS-CoV-2: implications for infection prevention precautions - Scientific brief*. Tech. rep. World Health Organization.

Covid Economics 53, 23 October 2020: 33-71

Annex

Description of the variables

The main variables are described as follows:

Dependent Variables

- **Number of weekly COVID-19 hospitalizations(Panel Data)** by race (Indigenous, Black, White, 'Pardos', East Asian people) at municipal level - From March 1^{st} to August 31^{th}, 2020.
- **Number COVID-19 infections (Covid-19 Cases SESAI)** for indigenous people recognized by the Special Secretariat for Indigenous Health (SESAI) within the Brazilian Unified Health System (SUS) - From April 1^{st} to August 31^{th}, 2020.
  - The obtained database counted the number of infections throughout the 34 Special Indigenous Health Districts (ISSD) located in the country. Since the analysis was executed at the municipal level, I needed to proportionally distribute the number of cases reported by the SESAI for each municipality that is located at least in one of these ISSDs (The distribution among ISSDs and municipalities can be found here).
  - I applied a *relative frequency* based on the estimates for the indigenous population in each of theses municipalities (IBGE, 2019) to find the corresponding proportion and, thus, determine the number of cases of infected from the ISSDs at the municipal level, as follow:

$$Sesai(x, y) = y_{i,d} \frac{x_i}{\sum_{n=j} x_j}$$

  - Where (y) is the daily number (d) of infected on a certain ISSD in which contains a determined municipality (i); (x) is the indigenous population of this particular municipality; and (j) is the set of all municipalities that are found on this specific ISSD.
- **Number of COVID-19 infections (Covid-19 Cases APIB)** for indigenous people recognized by the Articulation of the Indigenous Peoples of Brazil (APIB) - From April 1^{st} to August 31^{th}, 2020.

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

* The APIB granted us access to their database, which counted the daily number of the indigenous people infected not considered by the SESAI-SUS. As mentioned previously, the organization claims that the Brazilian Government (through SESAI) is under-notifying the actual number of indigenous infected by COVID-19. Although they gave us information at the municipal level, some instances did not have the municipality for the cases recognized and only the state in a respective day. Another point was that this database only accounted for a surplus of cases that SESAI did not consider in its reports.

* Similar to the estimated number of cases at the municipal level from the SESAI database (*Sesai(x,y)*), I needed to proportionally distribute the number of cases reported by the APIB for the instances that did not have municipalities, but only the states mentioned. Likewise, I did not account for all municipalities inside a state, but only those located inside in at least one Special Indigenous Health District (ISSD).

* I then applied a *Relative Frequency* based on the estimates for the indigenous population (IBGE, 2019) in each of theses municipalities to find the corresponding proportion and, thus, determine the number of cases of infected from the States at the municipal level, as follow:

$$Apib(x, y, z) = y_{i,d}\frac{x_i}{\sum_{n=j} x_j} + z_{i,d}\frac{x_i}{\sum_{n=s} x_s}$$

* Where (y) is the daily number (d) of infected on a certain ISSD in which contains a determined municipality (i); (x) is the indigenous population of this particular municipality; and (j) is the set of all municipalities that are found on this specific ISSD. Also, (z) is the daily number (d) of infected on a certain State in which contains a determined municipality (i); and (s) is the set of municipalities that are found on this specific State which are also located in at least one of the ISSDs.

- **Accumulated number of COVID-19 hospitalizations** by race (Indigenous, Black, White,'Pardos', East Asian people) at municipal level - From March 1^{st} to August 31^{th}, 2020. Data collected from DataSUS - SRAS database and retrieved at September 7^{th}, 2020.

Independent Variable of Interest

- **Deforestation**: Accumulated warning areas for deforestation (in 100 km^2) within the Amazon Forest and the Cerrado ecosystem at the municipal level - From March 1 to August 31, 2020. I also use the data in the natural log format. Data collected from Brazil's National Institute for Space Research - Deter database and retrieved at September 7 20^{th} 2020.
- **Deforestation (Panel data)**: Daily and weekly evolution to warning areas for deforestation (per 100 km^2) within the Amazon Forest and the Cerrado ecosystem at the municipal level - From March 1 to August 31, 2020.
- **Log deforestation**: Natural log of daily data of warning areas for deforestation (per km^2) within the Amazon Forest and the Cerrado ecosystem at the municipal level - From March 1 to August 31, 2020.

Other variables used as a control in the analysis at municipal level were:

1. Population data by race and total estimate of the Brazilian population. Data collected from Brazilian Institute of Geography and Statistics (IBGE) - 2010 Census and 2019 population estimates respectively;
    - **Indigenous population (% Total)**: Proportion of indigenous people at municipal level.
    - **'Pardos' population (% Total)**: Proportion of 'pardos' people at municipal level.
    - **Black population (% Total)**: Proportion of black people at municipal level.
    - **White population (% Total)**: Proportion of white people at municipal level.
    - **Yellow population (% Total)**: Proportion of yellow people at municipal level.
2. Access to multidisciplinary indigenous health care teams (EMSI). Data observed from March to June 2020 and retrieved at DataSUS - CNES Equipes de Saude.
    - **Average of multidisciplinary indigenous health care teams** outside and inside the Legal Amazon by municipality.
3. Data from the National Sanitation Information System (SNIS): (1) Water and Sewage [2018]; (2) Solid Waste [2018]; (3) Rainwater [2018];

Covid Economics 53, 23 October 2020: 33-71

- **Access to clean water**: Proportion of municipal population that have access to clean water in 2018.
- **Access to treated sewage**: Proportion of municipal population that have access to treated Sewage in 2018.
- **Access to public roads**: Extension of public roads inside the municipality (km) in 2018.
- **Urban density**: Proportion of urban population over the urban area of a municipality in 2018.
- **Urban density**: Total population of the municipality per Km^2.

4. Gini Index and GDP by municipality in Brazil. Data collected from SUS - Tabnet and Brazilian Institute of Geography and Statistics (IBGE), 2017 respectively;
   - **Inequality (Gini coefficient)**: Gini index based on GDP at municipal level in 2010.
   - **GDP**: Municipal GDP in 2017.
5. Cities with isolated indigenous populations. Data collected by the Instituto Socioambiental, COVID-19 and retrieved at August 14^{th}, 2020;
   - **Existence of uncontacted tribes**: Binary variable for municipalities that have uncontacted tribes (confirmed and not confirmed) in their territory in 2020. Source of the data: Instituto Socioambiental, 2020.
6. Geographical variables produced with the software GIS based on shapefiles from the Brazilian Institute of Geography and Statistics (IBGE);
   - **Rainfall**: Average of rainfall at the municipal level in millimeters per hour (mm).
   - **Waterway**: Binary variable for municipalities in which their centroids are at least 100 km distance to the nearest waterway.
   - **Distance to the coast**: Distance of the municipal capital until the closest cost area (km).
   - **Distance to the state capital**: Distance of the municipal capital until its state capital (km).
   - **Altitude**: Municipality altitude (m).

Covid Economics 53, 23 October 2020: 33-71

7. Illegal mining

   - **Illegal mining**: Binary variable for municipalities that have an illegal mining activity. The data was compiled by the Amazon Geo-Referenced Socio-Environmental Information Network (RAISG), Ilegal Mining Map and retrieved on August 14^{th} 2020.

8. Land conflicts;

   - **Number of land conflicts**: Total number of land conflicts involving Indigenous people in Brazil in 2019. Data collected by the 'Commissão Pastoral da Terra'.
   - **Land Conflicts - CACI**: Conflicts occurred involving indigenous people at the municipal level until 2019 (Dummy Variable). Created by the Fundação Rosa Luxemburgo, in partnership with Armazém Memória and InfoAmazonia, the database was collected at the Cartography of the Attacks against (CACI) website.

9. Cattle ranching;

   - **Cattle ranching**: The total number of bovine cattle by municipality per $1{,}000Km^2$. Also used in natural logarithmic form. Data from IBGE, *Censo Agropecuário* 2017.

10. Wildfire: Fire Radiative Power (FRP);

    - **Wildfire: Fire Radiative Power (FRP)**: Measurement of the radiant energy released per time unit by burning vegetation per $1{,}000Km^2$. Also used in natural logarithmic form. Data from INPE, Burning Program, for the Amazon and Cerrado regions.

11. Indigenous Territories

    - **Indigenous Territories**: Municipalities that have Indigenous lands officially recognized in Brazil (Updated in 2019). Data retrieved from IBGE - Indigenous and 'Quilombola' peoples database.

12. IBAMA - Environmental protection agency

    - **IBAMA**: Distance of centroid of a given municipalities to the nearest environmental protection agency (IBAMA) local office. The variable was calculated by the authors using GIS. Data retrieved from IBAMA.

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

Additional figures and tables

Figure 2: COVID-19 cases confirmed in Brazil, in the Developing and Developed countries

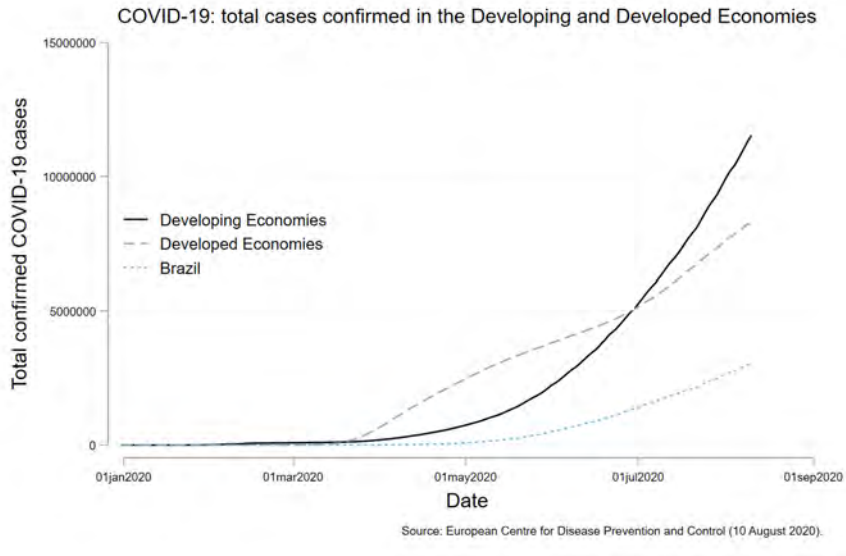


Table 6: Correlation matrix of selected (correlated) variables

Correlation matrix

	COVID-19 (SESAI)	COVID-19 (APIB)	COVID-19 hospitalizations	Deforestation	Wildfires	Cattle ranching	Illegal mining	Conflicts
COVID-19 (SESAI)	1							
COVID-19 (APIB)	0.9724	1						
COVID-19 hospitalizations	0.5241	0.4997	1					
Deforestation	0.2198	0.2408	0.1153	1				
Wildfires	0.2023	0.2294	0.1107	0.7618	1			
Cattle ranching	0.1406	0.1388	0.1257	0.3696	0.2927	1		
Illegal mining	0.3745	0.3837	0.2026	0.3709	0.3924	0.2228	1	
Conflicts	0.3170	0.3108	0.2039	0.2529	0.1688	0.2085	0.2319	1

Covid Economics 53, 23 October 2020: 33-71

Figure 3: COVID-19 cases in indigenous peoples and deforestation

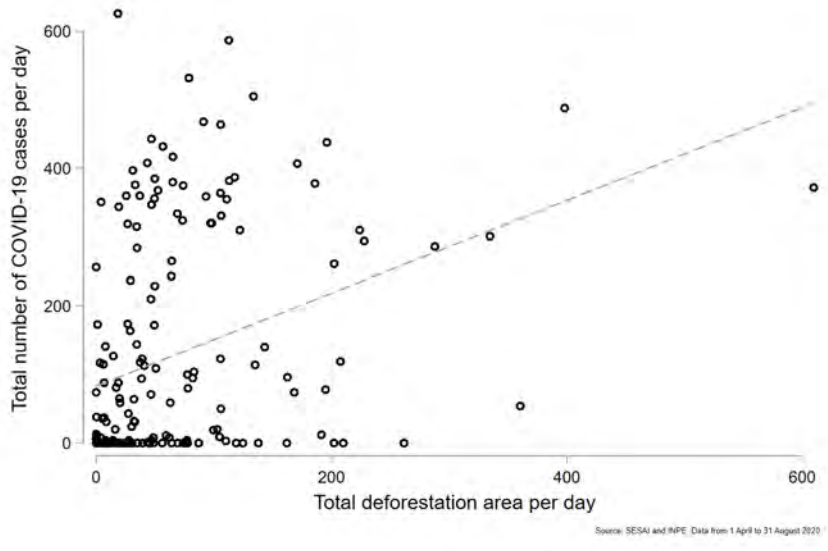


Table 7: Summary statistics: weekly panel data

Variables	N	Mean	SD	Min	Max	Sum
COVID-19 confirmed cases of indigenous people (SESAI)	146,260	0.158	2.299	0	136	23,179
COVID-19 confirmed cases of indigenous people (APIB)	146,260	0.198	3.188	0	294	28,985
COVID-19 hospitalizations of indigenous people	146,260	0.008	0.167	0	26	1,163
COVID-19 hospitalizations of black people	146,260	0.106	1.958	0	212	15,527
COVID-19 hospitalizations of white people	146,260	0.745	11.192	0	1,170	108,942
COVID-19 hospitalizations of 'pardo' people	146,260	0.761	9.117	0	738	111,290
COVID-19 hospitalizations of East Asian people	146,260	0.0257	0.527	0 [5]7	3,754	
Deforestation (per $100km^2$)	146,260	0.001	.0131	0	1.3	116.2

Covid Economics 53, 23 October 2020: 33-71

Table 8: Fixed-effects results: deforestation and COVID-19 cases in indigenous peoples

	COVID-19 cases SESAI				COVID-19 cases APIB			
	Pooled OLS	Fixed effects	Fixed effects	Fixed effects	Pooled OLS	Fixed effects	Fixed effects	Fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deforestation (per 100 Km^2)$_{t-5}$	3.247***	1.440**	4.739***		3.695***	1.641**	4.209***	
	(1.078)	(0.635)	(1.548)		(1.174)	(0.820)	(1.428)	
Log Deforestation (per Km^2)$_{t-5}$				0.0451**				0.0453*
				(0.0178)				(0.0260)
Deforestation (per 100 Km^2)$^2_{t-5}$			-9.446***				-7.350***	
			(2.933)				(2.580)	
COVID-19 cases SESAI$_{t-5}$	0.292***	0.164***	0.164***	0.100***				
	(0.0212)	(0.0166)	(0.0167)	(0.0294)				
COVID-19 cases APIB$_{t-5}$					0.138***	0.109***	0.109***	0.0431
					(0.0194)	(0.0220)	(0.0220)	(0.0459)
Implied cumulative	-0.699	-0.666	-0.569	-0.688	-0.877	-0.763	-0.687	-0.592
effect of deforestation	(0.577)	(0.562)	(0.518)	(0.561)	(0.651)	(0.639)	(0.600)	(0.631)
Constant	0.0168***	-0.00003	-0.00005	0.204***	0.0195***	-0.00008	-0.00009	0.285**
	(0.00198)	(0.00234)	(0.00234)	(0.0764)	(0.00222)	(0.00320)	(0.00320)	(0.113)
Observations	969,627	969,627	969,627	6,971	969,627	969,590	969,590	6,970
R-squared	0.083	0.030	0.030	0.067	0.048	0.014	0.014	0.080
Number of municipalities		5,417	5,417	913		5,417	5,417	913

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Weekly fixed-effects results: deforestation and COVID-19 cases in indigenous peoples

	COVID-19 cases SESAI				COVID-19 cases APIB			
	Fixed effects				Fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deforestation (per 100 Km^2)$_{t-1}$	6.370***				10.30***			
	(1.314)				(3.337)			
Deforestation (per 100 Km^2)$_{t-2}$		7.852***	25.61***			8.976***	33.53**	
		(2.293)	(8.420)			(2.999)	(13.03)	
Log deforestation (per 100 Km^2)$_{t-2}$				0.154**				0.300***
				(0.0668)				(0.110)
Deforestation (per 100 Km^2)$^2_{t-2}$			-32.97***				-45.59**	
			(12.31)				(19.33)	
Constant	-0.0003	-0.0004	-0.0013	0.671**	-0.0005	-0.0005	-0.0017	1.076***
	(0.0191)	(0.0199)	(0.0199)	(0.285)	(0.0245)	(0.0255)	(0.0256)	(0.299)
Observations	140,842	135,425	135,425	5,112	140,842	135,425	135,425	5,112
R-squared	0.009	0.009	0.012	0.044	0.007	0.006	0.009	0.033
Number of municipalities	5,417	5,417	5,417	914	5,417	5,417	5,417	914

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Covid Economics 53, 23 October 2020: 33-71

Table 10: State-fixed effects (accumulated): deforestation and COVID-19 cases in indigenous peoples

	COVID-19 cases SESAI				COVID-19 cases APIB			
	OLS	State fixed effects	State fixed effects	State fixed effects	OLS	State fixed effects	State fixed effects	State fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deforestation ($100km^2$)	41.59**	69.60***	69.31***	22.95***	59.25**	83.55***	80.84***	30.17***
	(15.18)	(10.88)	(4.067)	(4.728)	(25.21)	(12.26)	(5.579)	(5.931)
Population density	0.000221	0.00153*	0.000135	0.00128	0.000762	0.00284**	0.000412	0.00247
	(0.000700)	(0.000836)	(0.000595)	(0.00120)	(0.00107)	(0.00136)	(0.000817)	(0.00151)
Inequality (Gini coefficient)	103.5**	32.01**	21.28**	34.30***	131.0**	44.97**	33.10***	45.30***
	(43.09)	(12.82)	(8.328)	(10.06)	(50.26)	(17.04)	(11.42)	(12.62)
Extreme poverty			-1.645	-1.967			-2.493	-2.575
			(1.195)	(1.475)			(1.640)	(1.850)
Access to roads		1.042***	0.769***	0.0123		1.336***	0.913**	-0.0415
		(0.341)	(0.272)	(0.322)		(0.413)	(0.374)	(0.404)
Access to treated water		-4.393*	-1.520			-7.540**	-3.840	
		(2.434)	(2.355)			(3.619)	(3.230)	
Proximity to waterways		10.80	2.826	-0.00426		15.05	6.049*	0.757
		(6.706)	(2.478)	(2.772)		(9.140)	(3.399)	(3.477)
Number emergency rooms			6.749***	7.322***			12.01***	12.93***
			(0.979)	(1.452)			(1.343)	(1.821)
Constant	-48.64**	-19.80	-13.80***	-19.35***	-75.31**	-34.23	-20.05***	-26.71***
	(20.78)	(14.27)	(4.633)	(5.190)	(32.81)	(21.51)	(6.356)	(6.511)
Observations	5,380	2,414	2,413	3,261	5,380	2,414	2,413	3,261
R-squared	0.082	0.163	0.301	0.329	0.089	0.148	0.257	0.359
'Bad controls'	No	No	No	Yes	No	No	No	Yes
Geographical controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

CEPR PRESS

Table 11: State-fixed effects (accumulated): deforestation and mechanisms

	Log deforestation per $100Km^2$				
	OLS	State fixed effects	State fixed effects	State fixed effects	State fixed effects
	(1)	(2)	(3)	(4)	(5)
Log wildfires	0.0715***	0.0715***	0.0323***	0.0159*	0.0209**
	(0.00710)	(0.00568)	(0.00825)	(0.00842)	(0.00862)
Log cattle ranching	0.0425***	0.0457***	0.00225	-0.0117	-0.0123
	(0.00750)	(0.00694)	(0.00973)	(0.0106)	(0.0106)
Illegal mining	1.097***	1.027***	1.252***	1.234***	1.204***
	(0.348)	(0.121)	(0.229)	(0.228)	(0.228)
Conflict	0.658***	0.489***	0.545***	0.472***	0.488***
	(0.122)	(0.0582)	(0.0933)	(0.0939)	(0.0962)
Log local government			-0.0107	-0.00819	-0.00843
total revenues			(0.00951)	(0.00987)	(0.00988)
Distance to environmental protection			0.00744	-0.0336	-0.0382
agency's (IBAMA) nearest office			(0.0305)	(0.0338)	(0.0339)
Indigenous population/Total					-0.669
					(0.599)
Constant	-0.475***	-0.497***	0.144	0.0839	0.943
	(0.0673)	(0.0631)	(0.196)	(0.205)	(0.619)
Observations	5,385	5,385	1,801	1,698	1,698
R-squared	0.160	0.267	0.316	0.331	0.334
Geographical controls	No	No	No	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Covid Economics 53, 23 October 2020: 33-71

Table 12: Mechanisms: effects of deforestation on wildfires, cattle ranching, illegal mining, and conflicts

	Log wildfires		Log cattle ranching		Illegal Mining		Conflicts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log deforestation (km^2)	0.494***	0.407***	0.276***	0.166***	0.0159***	0.0148***	0.0326***	0.0321***
	(0.0318)	(0.0335)	(0.0265)	(0.0256)	(0.00148)	(0.00154)	(0.00307)	(0.00313)
Constant	1.706***	0.917***	9.141***	8.221***	0.00733***	-0.00778	0.0343***	0.0205**
	(0.0245)	(0.107)	(0.0203)	(0.0820)	(0.00114)	(0.00492)	(0.00236)	(0.0100)
Observations	5,417	4,994	5,417	4,994	5,385	4,994	5,417	4,994
R-squared	0.635	0.634	0.363	0.380	0.183	0.175	0.161	0.173
Geographical controls	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 13: Mechanisms: effects of wildfires, cattle ranching, illegal mining, and conflicts on the transmission of COVID-19 cases in indigenous people

	COVID-19 cases reported in indigenous people (SESAI)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log wildfires	1.028***	0.656**							0.318
	(0.257)	(0.276)							(0.257)
Log cattle ranching			0.444	-1.044***					-1.595***
			(1.062)	(0.386)					(0.360)
Illegal mining					153.7***	140.8***			129.3***
					(33.87)	(5.528)			(5.440)
Conflicts							61.15***	50.61***	42.58***
							(9.680)	(2.787)	(2.668)
Population density		0.00156**		0.000587		0.00137*		0.00142*	0.00002
		(0.000781)		(0.000846)		(0.000734)		(0.000756)	(0.000776)
GDP		-0.312		0.477		-0.574		-0.384	-0.994
		(1.058)		(1.033)		(0.971)		(0.999)	(0.972)
Inequality (Gini coefficient)		68.61***		74.02***		56.36***		61.47***	54.71***
		(9.268)		(9.334)		(8.714)		(8.966)	(8.596)
Proximity to waterways (100 km)		-3.263		-3.128		-0.300		-1.830	0.122
		(2.568)		(2.565)		(2.417)		(2.487)	(2.359)
Constant	2.486***	-38.25***	0.206	-39.03***	2.991***	-28.20***	2.032***	-34.42***	-10.97
	(0.648)	(11.24)	(10.09)	(11.12)	(0.995)	(10.35)	(0.325)	(10.64)	(10.54)
Observations	5,417	5,040	5,417	5,040	5,385	5,040	5,417	5,040	5,040
R-squared	0.123	0.120	0.001	0.120	0.148	0.220	0.098	0.173	0.260
Geographical controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Covid Economics 53, 23 October 2020: 33-71

Covid Economics 53, 23 October 2020: 33-71

Table 14: Mechanisms: effects of wildfires, cattle ranching, illegal mining, and conflicts on the transmission of COVID-19 cases in indigenous people

	COVID-19 cases reported in indigenous people (APIB)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log wildfires	1.581***	1.048***							0.520
	(0.334)	(0.361)							(0.336)
Log cattle ranching			0.735	-0.702					-1.448***
			(1.311)	(0.505)					(0.469)
Illegal mining					201.4***	189.5***			174.5***
					(48.93)	(7.207)			(7.102)
Conflicts							77.14***	65.78***	54.59***
							(12.59)	(3.649)	(3.483)
Population density		0.00261**		0.00187*		0.00233**		0.00239**	0.00112
		(0.00102)		(0.00111)		(0.000957)		(0.000990)	(0.00101)
GDP		0.528		1.588		0.319		0.605	-0.440
		(1.384)		(1.353)		(1.265)		(1.308)	(1.269)
Inequality (Gini coefficient)		87.96***		93.24***		71.92***		79.21***	67.44***
		(12.13)		(12.22)		(11.36)		(11.74)	(11.22)
Proximity to waterways (100 km)		-5.022		-4.682		-0.965		-3.079	-0.398
		(3.359)		(3.359)		(3.151)		(3.256)	(3.080)
Constant	2.594***	-59.22***	-1.385	-65.09***	3.662***	-47.14***	2.517***	-55.96***	-26.95*
	(0.842)	(14.70)	(12.09)	(14.56)	(1.208)	(13.49)	(0.413)	(13.94)	(13.76)
Observations	5,417	5,040	5,417	5,040	5,385	5,040	5,417	5,040	5,040
R-squared	0.101	0.107	0.001	0.106	0.153	0.214	0.095	0.160	0.252
Geographical controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 15: Cross-section: state fixed-effect estimates of the effect of deforestation on COVID-19 hospitalizations by race

	COVID-19 hospitalizations									
	Indigenous		Black		White		'Pardo'		East Asian	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Deforestation ($100km^2$)	0.825***	0.915***	-0.378	-0.596	-1.476	-3.380	13.27	14.87	0.210	0.0629
	(0.160)	(0.152)	(3.336)	(3.331)	(21.34)	(21.55)	(15.19)	(14.59)	(0.914)	(0.934)
Population density		0.0003***		0.0219***		0.127***		0.110***		0.0046***
		(0.00004)		(0.0008)		(0.0056)		(0.0038)		(0.0002)
PIB		0.000591		3.249***		23.43***		26.26***		0.930***
		(0.0524)		(1.150)		(7.440)		(5.037)		(0.323)
Inequality (Gini coefficient)		3.155***		54.07***		379.7***		259.6***		12.74***
		(0.469)		(10.30)		(66.63)		(45.11)		(2.889)
Constant	0.198***	-1.629***	2.875***	-57.31***	20.14***	-401.8***	20.28***	-371.7***	0.689***	-15.36***
	(0.0260)	(0.558)	(0.544)	(12.25)	(3.482)	(79.26)	(2.479)	(53.66)	(0.149)	(3.436)
Observations	5,417	5,040	5,417	5,040	5,417	5,040	5,417	5,040	5,417	5,040
R-squared	0.120	0.137	0.011	0.142	0.011	0.122	0.014	0.189	0.005	0.087
Geographical controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The triple effect of Covid-19 on Chinese exports: First evidence of the export supply, import demand and GVC contagion effects

Felix L. Friedt[1] and Kaichong Zhang[2]

Date submitted: 12 October 2020; Date accepted: 13 October 2020

In this study, we estimate the overall impact of the novel Coronavirus pandemic on Chinese exports and differentiate the hypothesized `triple pandemic effect' across its three components: 1) the domestic supply shock; 2) the international demand shock; and 3) the effects of global value chain (GVC) contagion. We find that Chinese exports are very sensitive to the severity of the global Coronavirus outbreaks. Average export elasticity estimates with respect to new Chinese and foreign destination country infections range from -2.5 to -4.6. Against a Covid-19-free counterfactual, our estimates predict that the pandemic has reduced Chinese exports by as much as 40% to 45% during the first half of 2020, but that these losses have peaked and are expected to partially recover by the end of the year. Moreover, we find that all three shocks contribute to the pandemic-induced reduction in Chinese exports, but that GVC contagion exerts the largest and most persistent influence explaining these losses. Among the three shocks, the impact of GVC contagion explains around 75% of the total reduction in Chinese exports, while the domestic supply shock in China accounts for around 10% to 15% and the international demand shock only explains around 5% to 10%. As a result of these varying transmission channels, the pandemic effects appear to be very distinct from those explaining the Great Trade Collapse in 2008-09.

1 Assistant Professor of Economics, Department of Economics, Macalester College.
2 Undergraduate economics and political science student, Macalester College.

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

1 Introduction

The novel Coronavirus has disrupted the global economy, and in particular international trade (Baldwin, 2020a). According to the 'World Trade Statistical Report 2020', published by the World Trade organization, global container throughput has dropped by 8% during the fist quarter of 2020, while new export orders declined by as much as 50% for both goods and services over the same time period (World Trade Organization, 2020a). Parallels have been drawn between the trade effects of the current pandemic and financial crisis of 2008, and some researchers anticipate the pandemic-induced contraction in international trade to overshadow the Great Trade Collapse of 2008-09 (Baldwin, 2009, 2020a). Baldwin and Freeman (2020), for example, argue that, unlike the financial crisis which stifled global demand for traded products, the pandemic triggers a 'triple effect' on trade through 1) the disruption of domestic supply, 2) the reduction in global demand, and 3) the contagion effect spread through disrupted global value chains (GVC). As the world's largest exporter and second largest importer of internationally traded goods and services and a central node in the GVC networks of a variety of products, China not only faces the ramifications of this global pandemic through its dependence on trade, but disruptions to Chinese exports may also be the root cause of the 'infection' of the globalized economy.

In this study, we estimate the overall impact of the Coronavirus pandemic on Chinese exports and explore the heterogeneity of trade effects across Chinese provinces, international trade partners, and commodities. Moreover, we dissect the hypothesized 'triple pandemic effect' on trade and evaluate the individual contributions of the pandemic-induced domestic supply, international demand, and GVC contagion shocks. Our investigation sheds first light on the sensitivity of international trade flows to the pandemic along all three of these transmission channels and shows that Chinese exports not only fall in response to a rise in domestic and international destination country Covid-19 cases, but also through increasing exposure of Chinese production to the GVC contagion. Our baseline empirical analysis of the elasticity of Chinese exports with respect to domestic and foreign infections provides economically and statistically significant estimates suggesting that for every 1% rise in new domestic and international Coronavirus cases Chinese exports fall by 2.5% to 4.6%. In aggregate, these estimates suggest that the short-run losses in Chinese exports during the first half of 2020 may be as large as 40% to 45% of the predicted counterfactual Chinese exports in the absence of the pandemic.

The differentiation of the individual transmission mechanisms shows that all three channels contribute

Covid Economics 53, 23 October 2020: 72-109

to the reduction in Chinese exports, but do so at different points in time and to varying degrees. The international demand effect, captured by the number of new foreign Covid-19 cases, for example, has an immediate negative impact on Chinese exports that lasts for about one month after the rise in new cases. Thereafter, it appears that the disruption of their own production causes foreign countries to increase their demand for imports from China, which in turn offsets some of the initial demand-driven losses. The domestic (Chinese) supply effect on Chinese exports, captured by the number of new Chinese Coronavirus cases, is slightly lagged and strongest one to two months after a rise in local cases. At their respective peaks, both of the pandemic-induced domestic supply and international demand shocks are similar in magnitude and account for about 30% to 40% of the reduction in Chinese exports. The GVC contagion effect on Chinese exports is significant and delayed by about one month. That is, one month after a rise in the global Covid-19 exposure of intermediate inputs that are used in production of Chinese exports, the GVC contagion takes a negative toll. We find that these ripple effects are significant and account for the majority of the reduction in Chinese exports. In aggregate, more than 75% of the total pandemic effect on Chinese exports can be explained by the disruption of GVCs, which dissipates much slower than the domestic supply and international demand shocks.

Lastly, we take advantage of the disaggregated Chinese customs data and explore idiosyncrasies in the Covid-19 effects along three distinct dimensions, including Chinese provinces, foreign countries, and commodities. Several notable patterns emerge. First, not all provinces suffer the same fate in terms of their exports. While the majority of Chinese regions experience the expected significant decline in exports, a handful of provinces report year-over-year increases in exports between the first halves of 2019 and 2020 (i.e. Anhui, Bejing, Jiangxi, etc.). Second, the sensitivity of Chinese exports with respect to foreign cases seems to decline with a country's level of development and distance to China, perhaps indicating that higher income countries are more dependent on Chinese exports and/or better equipped to continue the flow of international goods during this pandemic. Lastly, we find that the nature of this global pandemic and the unique regulatory responses shape the impact on Chinese exports of various commodities. While exports of pharmaceutical products significantly rise with the intensity of the Coronavirus outbreak, products that are typically air freighted suffer significant reductions in exports due to the collapse of international air travel.

To derive at these conclusion, we develop a new dataset that combines detailed provincial customs data on Chinese exports from January 2019 through June 2020 with information on the spread of the novel

Coronavirus both in China and overseas. Furthermore, we augment these data with country-industry-specific statistics on 'Foreign Value Added' and 'Indirect Value-Added Exports' (Koopman et al., 2014) that are derived from national and regional input-output tables and part of the UNCTAD-Eora Global Value Chain Database (Casella et al., 2019). Combining Chinese and international Coronavirus case counts with information on value-added trade, we are able to characterize the industry-specific exposure of GVCs pertinent to Chinese production and evaluate the resulting GVC contagion effect on Chinese exports in isolation from the domestic Chinese supply and international demand shocks.

Our work builds on the previous research investigating the trade effects of other major disruptions, such as the Great Trade Collapse of 2008-09 or the Great Japanese Earthquake of 2011, and contributes to the rapidly growing literature on the economic effects of the global Coronavirus pandemic. Much of the very recent research has focused on the pandemic-induced demand and supply shocks in a variety of different contexts, such as the U.S. stock and labor markets (see, for example, Papanikolaou and Schmidt (2020) and Baker et al. (2020), or Cajner et al. (2020) and Coibion et al. (2020), among others). In short, the supply shock is attributed to the public-health-related containment measures that severely limit the mobility and productivity of workers and the demand shock is the result of the unprecedented spike in unemployment (Coibion et al., 2020), reductions in consumer spending (Bachas et al., 2020) and investment uncertainty that leads to a 'wait-and-see' mode (Baldwin, 2020b).

Some research has considered the pandemic effects on international trade. Bonadio et al. (2020), for example, offer a general view of the pandemic shock on international trade. The authors evaluate the resilience of economies to pandemic-induced contractions, suggesting that the economic destruction is mainly caused by the stringent containment measures and the reopening of the economy, especially by the largest economies like US, China, Russia, Germany, and Japan, would generate a considerable positive impact (Bonadio et al., 2020). Leibovici et al. (2020) evaluate the pandemic effects on trade of essential goods and find that the ensuing welfare effects critically depend on a country's trade balance with respect to those commodities. Drawing parallels between the Coronavirus pandemic and the SARS outbreak in 2003, Fernandes and Tang (2020) estimate the impact of SARS on firm-level Chinese trade to gain insights into the likely effects of the current pandemic. Fernandes and Tang (2020) find that SARS impacted Chinese producers for over two years after the initial outbreak and that the effects varied across firms. Small firms are found to be more likely to exit, while producers of more capital- and skill-intensive commodities or those that are more upstream on the supply chain drive the recovery of Chinese trade post SARS.

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

Baldwin and Freeman (2020) as well as Baldwin and Tomiura (2020) develop a conceptual framework that closely considers the specific mechanisms driving these trade effects. The authors argue that the pandemic has the potential to cause a 'triple hit' on manufacturing sectors, including the direct supply disruption due to various containment efforts, the supply-chain contagion due to the disruptions of the international flow of intermediate inputs, and the decline in global demand due to reduction in consumer spending and investment delays. Using the textile as well as information and communication technology (ICT) industries as examples, Baldwin and Tomiura (2020) also point to China as the 'workshop of the world' central to a multitude of GVCs and critical to understanding the comprehensive pandemic effect on international trade. We contribute to the literature by providing first estimates of the pandemic effects on Chinese exports along all three of these dimensions and offering initial insights on the significance of the theorized GVC contagion.

As such our work is closely related to the analyses of other large-scale economic shocks on international trade, such as the Great Trade Collapse 2008-09 (GTC) or the 2011 Japanese Earthquake. Previous research on the GTC of 2008-09, for example, has shown that the past demise in trade outpaced the decline in global GDP and can be largely attributed to the reduction in global demand (Crowley and Luo, 2011; Levchenko et al., 2010; Altomonte et al., 2012) as well as constraints on trade finance during the Great Recession (Ahn et al., 2011; Chor and Manova, 2012).[1] Baldwin (2020a) has drawn parallels between the GTC and the Coronavirus pandemic and argues that the latter trade shock may be more intense due to the multifaceted supply and demand transmission channels we study here.

In contrast to the demand-driven GTC, the 2011 Japanese Earthquake impacted trade through the supply disruption that resulted from significant damage to production, energy and transportation infrastructure in Japan. While Ando and Kimura (2012) show that the domestic supply disruptions reduced Japanese trade, research by Boehm et al. (2019) and Li et al. (2015) illustrates that even such a unilateral shock can cause significant international spillover effects due to the interconnectedness of global production networks. One of the critical distinctions between the Japanese Earthquake of 2011 and the current Covid-19 pandemic, however, is one of scope. While the former disaster hit a single hub in the global supply chain

[1] Levchenko et al. (2010), for example, show that the 2008-09 GTC exceeded the reduction in global GDP due to the financial crisis and argue that the amplified response in trade is a result of its composition. The authors show that the majority of trade is comprised of durable goods which tend to be more sensitive to negative demand shocks than non-durables. (Altomonte et al., 2012) advance this discussion and suggest that a reduction in demand for final goods can lead to an amplified response in the demand for traded intermediate products further up in the supply chain. The depletion of inventories, strategically build to offset uncertainties in downstream demand, has a 'bullwhip' effect that propagates and amplifies the negative demand shock throughout the GVC network and is the counterpart to the GVC contagion we investigate in this study.

network, the latter has impacted all of the major manufacturing hubs around the globe (Baldwin and Freeman, 2020). The ability of affected firms to find qualified substitutes for intermediate inputs from other unaffected countries contributed to the rapid recovery from the 2011 Japanese Earthquake (Todo et al., 2015), but is a luxury that is not necessarily afforded by the Covid-19 pandemic. Our estimates produce evidence to suggest that, while the domestic supply and international demand shocks are economically and statistically significant, the GVC contagion effect explains the majority of the disruption in trade and causes longer lasting ramifications for Chinese exports.

The remainder of the paper is organized as follows. Section 2 provides background information on the novel Coronavirus outbreak and regulatory responses in China and the rest of the world, and offers a first glimpse at the trade effects resulting from the pandemic. In section 3, we develop a conceptual framework that is informed by the discussion of Baldwin and Tomiura (2020) and derive our baseline and primary empirical specifications. In section 4, we discuss the construction of our dataset and summarize the relevant statistics. We present our empirical results in Section 5. Lastly, we provide a brief discussion of our analysis in section 6 and conclude in section 7.

2 Background

The first cases of Covid-19 were reported in Wuhan China dating back to November 17, 2019 according to the South China Morning Post (Ma, 2020). Accelerated by the mass holiday travel during Chunyun (the Spring Festival travel season), the virus spread rapidly in the Hubei province and throughout China. Figure 1 illustrates the steep ascent of new confirmed Chinese Coronavirus infections in January and February, 2020. Throughout February, for example, new confirmed cases in China increased by 572% from 11,821 to 79,389 confirmed infections. As a result, China became the first epidemic hot spot in 2020.

In order to slow the spread of the infection, the Chinese government elected rigorous quarantine policies locking down entire metropolitan cities with millions of residents, such as Wuhan, Qianjiang, and Huanggang in the Hubei province. Furthermore, regulators decided to block major highways, railways, and air lines between cities. Local municipalities followed suit and issued similar policies, such as the prohibition of gatherings (even between family members) and ceasing all public transportation, in order to curb the spread of the novel Coronavirus. In some cases, locally issued permits allowed essential activities, such as grocery shopping, but were restricted to the district of residence. In the most infected areas, such as Wuhan,

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

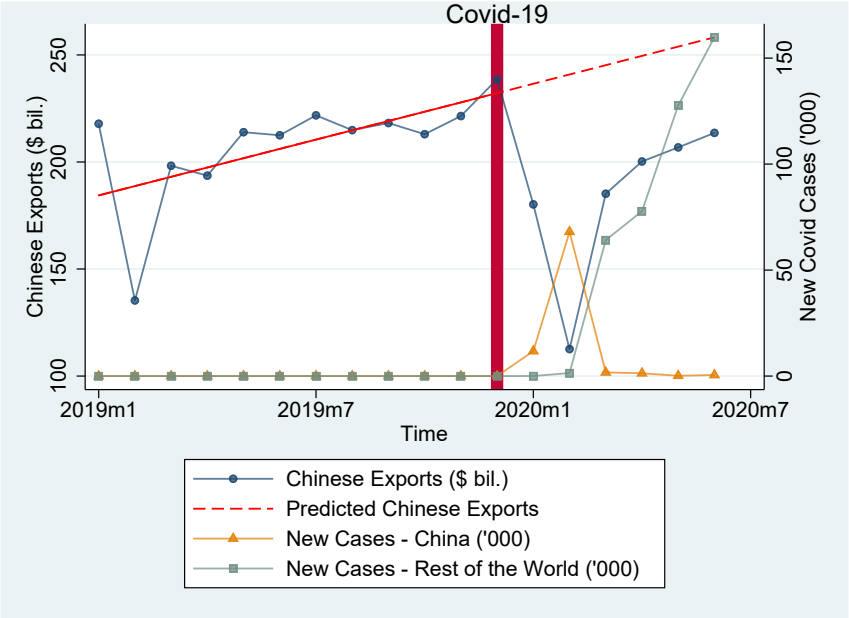


Figure 1: Chinese Exports and Covid-19 Cases

however, even grocery shopping was banned by the government and replaced by a centralized grocery delivery system to provide residents with daily necessities. Under such strict control, China quickly reached the peak of new Coronavirus infections, but also severely restricted economic activity. New daily confirmed cases in all of China dropped from 15,153 on February 17 to 206 by March 1, 2020, while a Chinese production index, published as part of the Federal Economic Reserve Database, suggests an average contraction in manufacturing sectors of over 25% during this time period.

Since then, the Chinese government has shifted its focus toward economic recovery. In terms of fiscal policy, the central government of China announced a 4.6 trillion Yuan stimulus package and synchronized financial support for manufacturers of medical and daily necessities as well as the agricultural sector. By March and April, many cities enacted plans to reopen the economy and issued guidance on returning to work according to the number of new local cases. Mobility and activity restrictions were removed gradually and prioritized the return of workers to essential industries (International Monetary Fund, 2020). During this process, any detection of new infections, such as those experienced by Shulan Jilin in May 2020, triggered an immediate emergency status restricting residential mobility and placing residential districts under

Covid Economics 53, 23 October 2020: 72-109

strict surveillance. Most cities, however, successfully reopened their economies and China experienced a 3.2% real GDP growth rebound by June 2020.

While China experienced first signs of recovery during the second quarter of 2020, the novel Coronavirus began to spread throughout the rest of the world. Europe became the second hot spot of the pandemic, followed by the United States, which surpassed China and Italy in terms of total confirmed deaths by mid April and remains the country with the most Coronavirus infections today. Figure 1 documents this rapid increase in new international infections and shows no signs of reprieve by June 2020. In fact, by August 2020, the outbreak of Covid-19 has affected 188 countries with 18 million confirmed cases and 700,000 confirmed deaths; and on September 1, the number of global daily new cases reached 439,370, the highest point since the outbreak of Covid-19.

With the rise of new infections, many governments enacted a number of different policies in an attempt to limit further contagion. Most European countries, such as Italy or Germany, adopted severe containment policies ranging from travel restrictions to bans on every form of public gathering. The United States also declared a state of emergency, increased testing, closed schools and non-essential businesses, and implemented social distancing to fight the spread of the pandemic (International Monetary Fund, 2020). Many of these policies were enforced until late May and early June, when most of the European countries and the United States announced the relaxation of Covid-related restrictions allowing businesses to reopen and travel to resume among Schengen-area members. In turn, a second wave of Coronavirus infections emerged shortly thereafter and by July 17 daily new cases in the United States reached the peak with 75,821 new confirmed cases.

Over this time period, the world economy experienced one of the most severe contractions in recent history. Early research efforts estimate a 2% to 4% reduction in global GDP due to Covid-19 (Maliszewska et al., 2020), with some countries expected to experience economic contractions of more than 10% under pessimistic forecasting assumptions (Fernandes and Tang, 2020). Similar to the GTC in 2008-09, the response in international trade is expected to surpass these reductions in GDP and some experts anticipate that the collapse in international trade due to the pandemic will exceed the GTC in 2008-09 (Baldwin, 2020a). Early estimates by the WTO, for example, predict global trade in 2020 to decline by as much as 32%. Barichello (2020), who analyzes Canadian agricultural trade, posits an expected decline of 12% to 20% despite the fact that agricultural commodities are relatively insulated from changes in demand. Our data presented in Figure 1 offer some preliminary evidence that corroborates these projections. A comparison between Chinese

Covid Economics 53, 23 October 2020: 72-109

exports in December 2019 and February 2020 suggests a decline of $126 billion or 53%. These statistics, however, likely overstate the immediate effect of Covid-19 as the time period under consideration includes the Chinese New Year celebrations, during which exports tend to fall significantly (see February 2019 in Figure 1). The year-over-year change in exports suggests a decline of $37 billion in January and $23 billion in February and equates to a 17% reduction in both months relative to the previous year. This year-over-year comparison, however, ignores the substantial growth Chinese exports experienced in 2019 and therefore underestimates the true pandemic-induced trade effect. Combining the year-over-year reduction with the anticipated export growth, that is forgone due to the pandemic, yields a rough estimate of Chinese export losses as high as 42% in January and February and around 30% to 35% thereafter.

3 Conceptual Framework

The gravity equation of international trade is the empirical 'workhorse model' for trade economists (Head and Mayer, 2014) and a good point of departure for evaluating the potential trade effects of the Coronavirus pandemic. In the traditional framework, exports (X_{ij}) from country i to country j are positively correlated with the economic mass or GDP (Y) of both countries and inversely related to the distance and other barriers to trade, such as tariffs, between them (τ_{ij}). In its simplest form the gravity model of trade can be expressed as follows:

$$X_{ij} = \frac{Y_i Y_j}{\tau_{ij}}. \tag{1}$$

And, several necessary amendments have been made in previous research to account for the complexities of a global and interconnected economy that influence international trade flows (see, for example, Anderson and Van Wincoop (2003)).

Nonetheless, this simple model lends itself to conceptualize some of the potential mechanisms through which Covid-19 can influence international trade. On the exporter's side, the outbreak of the novel Coronavirus can disrupt the production of tradable goods and services. In fact, mass quarantines and entire city lock-downs as coordinated and mandated in China, for example, have led to a temporary standstill in the production of the world's largest supplier of manufactured goods. In the simple gravity equation (1), this adverse domestic supply shock is reflected in a reduction of country i's GDP, which in turn reduces country i's exports.

Covid Economics 53, 23 October 2020: 72-109

On the importer's side, the impact of a rise in infections with the novel Coronavirus is more ambiguous. On the one hand, more infections can result in a disruption of local production which may result in a substitution away from domestic goods and towards imported products, which would stimulate exports to the affected destination country. On the other hand, the likely reduction in consumer demand, investment uncertainty, and disruption of local production that uses foreign intermediate inputs can drastically reduce the demand for imported products. An example of a such a shock is given by the GTC of 2008-09, which has been largely attributed to the reduction in global demand (Crowley and Luo, 2011).[2] Moreover, in terms of the anticipated effect of a reduction in demand, several factors may amplify the impact on internationally sourced, rather than domestically produced, goods. First, the majority of trade is comprised of durable goods which have been shown to be more sensitive to negative demand shocks than non-durables (Levchenko et al., 2010). Second, a reduction in demand for final goods can lead to an amplified response in the demand for traded intermediate products further up in the supply chain. The depletion of inventories, strategically build to offset uncertainties in downstream demand, has a 'bullwhip' effect that propagates and amplifies the negative demand shock throughout the GVC (Altomonte et al., 2012). In our gravity equation, this adverse international demand shock is reflected in a reduction of the importing country j's GDP, which in turn reduces country i's exports to country j.

In addition to these first order shocks to the supply of the exporting country and the demand of the importing country, the pandemic has the potential to cause notable ripple effects through GVCs. Unlike the Great Trade Collapse of 2008-09, these ripple effects are rooted in the significant disruption of the production of intermediate inputs experienced by the major suppliers around the globe. Coining this the 'GVC contagion' effect, Baldwin and Tomiura (2020) argue that the early epicenters of the Coronavirus outbreak, including China, U.S., Japan, Germany, and South Korea, are also manufacturing hubs for a large number of inputs used by numerous GVCs spread acorss a broad network of countries. The interconnectedness among these hubs likely amplifies local disruptions and spreads the economic pandemic effects to many other countries reliant on these centers of production.

In order to gain some initial insights on all three of these pandemic effects on international trade, we develop an empirical framework of Chinese provincial-level exports that captures each of these potential transmission channels. Given the recency of the Coronavirus outbreak and lag in the reported statistics on

[2]There is some evidence that constraints on the trade finance during the Great Recession also played a role in the demise of trade during this time period (see, for example, or Ahn et al. (2011) or Chor and Manova (2012)).

GDP, however, we must rely on an alternative measure that (1) captures the intensity of the pandemic and (2) strongly correlates with the potential shocks to domestic supply and international demand. We argue that case counts of new Coronavirus infections are one such measure giving a timely and location-specific insight on the severity of the pandemic. Moreover, we believe that this pandemic statistic is indicative of the economic strains and regulatory restrictions placed on an economy to curb further contagion of the virus and therefore correlated with the potential supply and demand shocks on international trade.

Accordingly, we model exports (X_{pjkt}) from Chinese province p to foreign country j of commodity k at time t as a function of the number of new Coronavirus infections reported by Chinese provinces (CI_{pt}) and their foreign trade partners (CI_{jt}). While provincial-level pandemic statistics are indicative of the domestic export supply shock, international case counts should provide a window into the expected foreign import demand shock. The baseline empirical specification then takes the following form:

$$x_{pjkt} = \beta_0 + \beta_1 ci_{pt} + \beta_2 ci_{jt} + \alpha_{pj} + \alpha_k + \alpha_t + \epsilon_{pjkt}, \tag{2}$$

where the lower case letters represent the inverse hyperbolic sine (IHS) transformation of the aforementioned variables.[3] In line with many empirical specifications of the gravity equation (Head and Mayer, 2014), we control for province-to-country (α_{pj}), commodity (α_k), and time (α_t) fixed effects that are intended to capture differences in time-invariant bilateral trade costs, such as distance, industry characteristics, i.e. durable versus non-durable products, and macroeconomic trends influencing Chinese exports across all countries and provinces.

Because we intend to estimate the effects of domestic and foreign Coronavirus infections on Chinese exports, we cannot separately include time-varying country- or province-specific fixed effects. Consequently, a limitation of this specification is the inability to control for time-varying country- or province-specific unobservables, such as changes in bilateral trade costs during our sample period. If such an unobservable correlates with the outbreak and spread of Coronavirus infections, either in China or overseas, our estimates of β_1 and/or β_2 would be biased. One particular concern is the ongoing U.S.-China trade war. To mitigate the influence of these contemporaneous commercial policy changes we exclude the U.S. from our primary estimation sample and test sensitivity of our results against this restriction. Our findings are

[3] We use the IHS transformation, rather than the natural log, because it is defined at zero-valued trade observations and infection counts and yields coefficient estimates that are readily interpretable as export elasticities. The results are generally robust to the logged transformation and exclusion of zero-valued observations.

Covid Economics 53, 23 October 2020: 72-109

largely robust to this adjustment.

To capture the third channel of the pandemic effect, namely the GVC contagion, we integrate an additional variable in our baseline specification. The intent is to create a measure that captures the exposure of Chinese export industries to the 'infection' of the underlying GVC networks these industries depend upon. This 'infection', of course, is a result of the Coronavirus-induced disruptions of foreign suppliers. To this end, we calculate the monthly share of global infections each foreign country faces. We then weight these relative infection shares by the share of foreign value added (FVA) a country provides to Chinese exports of a particular commodity.[4] In other words, each share provides a snap shot of how 'infected' a particular link of the supply chain network is, while the weights measure how important this link is to Chinese exports of a particular commodity.[5] Suppose, for example, that Germany is the sole supplier of foreign inputs to Chinese export industry A, but supplies no intermediate inputs for Chinese export industry B. If Germany is the only country affected by the novel Coronavirus and reports 100% of new global infections, than Chinese export industry A faces the maximum degree of GVC contagion, while export industry B faces none.

In reality, GVCs are, of course, much more complex involving many countries at different stages of production. To capture the level of 'infection' of such complex supply chains, we sum these weighted shares across the full set of foreign countries J. Aggregated, our variable gives us a sense of the total exposure that Chinese industries of exported products face due to the pandemic-induced disruption of GVCs. The exact expression for our measure of GVC contagion (GVC-C) is given as follows:

$$\text{GVC-C}_{kt} = \sum_{j=1}^{J} \left[\frac{FVA_{jk\bar{t}}}{\sum_j FVA_{jk\bar{t}}} * 100 * \frac{CI_{jt}}{\sum_j CI_{jt}} * 100 \right], \tag{3}$$

where we fix the level of FVA in 2016 prior to the Coronavirus outbreak (and U.S.-China trade war) to avoid any contemporaneous feedback effects on the GVC network characterized by FVA.

Finally, we argue that the timing of the domestic supply shock, international demand shock, and GVC contagion on Chinese exports may vary for each of these transmission channels. That is, one may expect the domestic supply and international demand shocks, captured by rising Chinese and foreign destination country Coronavirus infections, to take effect before the GVC contagion, which must first ripple through the

[4]FVA is a term often used in the GVC literature (see, for example, Johnson and Noguera (2012) and Koopman et al. (2014)) and measures the amount of the total value of Chinese exports that is attributable to intermediate inputs produced in foreign countries, exported to China, and then used in the production of Chinese exports.

[5]By the level of 'infection' we mean the severity of the Covid-19 outbreak in a given country relative to all other Chinese trade partners.

Covid Economics 53, 23 October 2020: 72-109

CEPR PRESS

GVC network. Consequently, our final and preferred empirical specification includes the contemporaneous ($s = 0$) as well as one- through three-months ($S = 3$) lagged terms of our key variables and can be expressed as follows:

$$x_{pjkt} = \beta_0 + \sum_{s=0}^{3} \left[\beta_s ci_{pt-s} + \lambda_s ci_{jt-s} + \psi_s \text{GVC-C}_{kt-s}\right] + \alpha_{pj} + \alpha_k + \alpha_t + \epsilon_{pjkt}. \quad (4)$$

The coefficients of interest can be collected in three distinct sets. Coefficient vector β captures the contemporaneous and lagged domestic supply shock on Chinese exports. Coefficient vector λ captures the contemporaneous and lagged international import demand shock on Chinese exports. Lastly, coefficient vector ψ measures the contemporaneous and lagged effects of the GVC contagion that may arise as the supply of intermediate inputs from foreign countries is disrupted by the pandemic.

4 Data

To estimate Equations (2) and (4) and test whether any of these three transmission channels influence trade during the time of the Coronavirus pandemic, we construct a new dataset that combines Chinese customs data, sourced from the General Administration of Customs of the People's Republic of China (GACC), with detailed information on the spread of Covid-19, both in China and a number of foreign countries. The GACC trade data report exports at the Chinese Province - Foreign Country - two-digit HS commodity level on a monthly frequency. The full sample includes exports of 97 commodity classes originating from 31 Chinese provinces that are destined for 241 foreign countries and runs from January 2019 to June 2020.[6] We combine these trade data with detailed information on the Chinese Coronavirus outbreak published in the monthly statistical report of China's National Health Commission. Staring in January 2020, the data include the new confirmed case and death counts observed in each province and provide insight into the severity of each province's Covid-19 supply shock.

International Covid-19 data at the country level are publicly available from the European Centre for Disease Prevention and Control (ECDC) and include monthly case and death counts for many foreign countries, the first of which are reported in January 2020. Because we don't observe international infection and death counts for every country, a merger between the ECDC and GACC datasets reduces the number

[6] Due to delays resulting from the Coronavirus outbreak, Chinese authorities opted to combine exports in January and February 2020 as a single month. We split these data according to the January and February shares in 2019. This essentially smooths the immediate impact across the first two months. Our results are robust to alternative strategies for allocating the combined statistics to the individual months.

Covid Economics 53, 23 October 2020: 72-109

of foreign nations to 199. Furthermore, it is important to note that case and death counts are based on daily reports and subject to some mismeasurement due long reporting chains. To address this issue we use aggregated monthly statistics and restrict the primary estimation sample to a set of 64 countries for which we observe case and death counts for six consecutive months in 2020. We test the sensitivity of all of our results against this sample restriction and obtain robust coefficient estimates when we expand the set of countries.

Lastly, we augment our trade and Covid-19 data with detailed information on country-indstury-specific value added trade. Specifically, we derive data on FVA contributing to Chinese exports and 'indirect value-added exports' (DVX) provided by Chinese exporters to foreign country trade from the UNCTAD-Eora Global Value Chain Database.[7] These data are based on a host of national and regional input-output tables and can be used to characterize the global GVC network of over 187 countries and 123 industries from 1990 to 2019 (2017-2019 values are forecasted) (Casella et al., 2019). To avoid any contemporaneous feedback effects from the pandemic and forecasting errors for the most recent years, we hold these statistics constant at their respective 2016 values. To finalize our data, we build an industry-commodity concordance that maps 84 of the 123 Chinese sectors reported in the EORA dataset and 86 out of 97 traded commodities into 43 categories of industries producing tradable goods. Combining the information on FVA with the international cases counts and Chinese trade data allows us to calculate our GVC contagion measure for 43 export industries according to Equation (3).

We summarize this novel dataset by considering each of its unique dimensions separately. In Table 1, we begin from the perspective of Chinese provinces. Column (1) provides a concordance between Chinese provinces and regions. whereas columns (2) through (7) offer a detailed view on the provincial exposure to the domestic supply shock, the international demand shock, and the GVC contagion. Columns (2) through (5), for example, report monthly averages as well as cumulative totals of provincial case and death counts from January through June 2020. The data show a large degree of variation in the severity of the provincial-level outbreaks with total confirmed cases ranging from 1 in Tainjin to 68,135 in Hubei, where the city of Wuhan is located. Consequently, we expect the intensity of the domestic supply shock to also vary significantly across these provinces.

[7] FVA and DVX are measures borrowed from the GVC literature. At the bilateral level, these represent the individual country-industry-to-country-industry linkages that make up the GVC networks. While FVA measures the importance of foreign country exports to Chinese producers of exported commodities, DVX determines how important industry-specific Chinese exports are as an intermediate input in the value added that is exported by foreign countries (Koopman et al., 2014).

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

Table 1: Chinese Provincial Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Export Exposure to Covid-19							Yr.-over-Yr. Change (Jan.-Jun. 2019-20)	
Chinese Province	Chinese Region	Avg. New Cases per Month	Avg. New Deaths per Month	Total Cases	Total Deaths	Trade wgt. Foreign Cases	GVC Cont.	Total Exports ($ bil.) (Jan.-Jun. 2019)	Exports ($ Bil.)	Exports (%)
Anhui Province	East	165.2	1.0	991	6	10,104.6	64.7	19.2	0.6	3.2
Beijing	North	153.7	1.5	922	9	3,637.0	57.0	35.5	0.1	0.2
Chongqing	Southwest	97.0	1.0	582	6	11,688.5	67.7	25.2	-1.0	-4.0
Fujian Province	East	60.5	0.2	363	1	9,010.1	61.6	59.9	-7.8	-13.1
Gansu Province	Northwest	27.3	0.3	164	2	3,257.6	60.0	1.0	-0.4	-43.0
Guangdong Province	South Central	273.5	1.3	1,641	8	8,066.6	62.6	292.0	-33.7	-11.5
Guangxi Zhuang Aut. Region	South Central	42.3	0.3	254	2	4,130.3	61.8	19.3	-2.4	-12.2
Guizhou Province	Southwest	24.5	0.3	147	2	3,660.3	63.2	2.1	0.1	6.5
Hainan Province	South Central	28.5	1.0	171	6	5,568.7	56.3	2.4	-0.4	-18.2
Hebei Province	North	58.2	1.0	349	6	7,763.1	62.8	16.0	-0.3	-1.6
Heilongjiang Province	Northeast	39.7	0.2	238	1	5,388.8	57.9	2.4	0.0	0.6
Henan Province	South Central	212.7	3.7	1,276	22	14,613.9	62.7	21.3	-0.4	-2.1
Hubei Province	South Central	11,355.8	752.0	68,135	4,512	8,615.4	63.7	15.7	-1.7	-10.7
Hunan Province	South Central	169.8	0.7	1019	4	5,785.9	61.1	18.9	0.5	2.5
Inner Mongolia Aut. Region	North	157.8	2.2	947	13	5,074.3	53.2	2.7	-0.3	-11.8
Jiangsu Province	East	109.0	0.0	654	0	9,805.0	64.2	189.9	-16.5	-8.7
Jiangxi Province	East	155.3	0.2	932	1	7,149.9	63.4	16.8	3.6	21.5
Jilin Province	Northeast	33.0	0.0	198	0	5,178.4	60.1	2.4	-0.3	-11.7
Liaoning Province	Northeast	118.7	1.2	712	7	5,510.5	63.0	22.8	-4.4	-19.3
Ningxia Hui Aut. Region	Northwest	12.5	0.0	75	0	7,413.7	54.0	1.1	-0.5	-45.0
Qinghai Province	Northwest	3.0	0.0	18	0	5,346.2	46.3	0.2	-0.1	-64.6
Shaanxi Province	Northwest	53.3	0.5	320	3	4,848.4	65.0	14.3	-1.7	-11.9
Shandong Province	East	132.0	1.2	792	7	7,704.1	62.4	77.6	-2.0	-2.6
Shanghai Province	East	25.8	0.3	155	2	9,686.1	63.9	93.9	-2.4	-2.5
Shanxi Province	North	25.8	0.3	155	2	16,000.1	63.1	5.1	-0.6	-12.5
Sichuan Province	Southwest	99.2	0.5	595	3	10,295.8	65.5	24.8	3.8	15.4
Tianjin	North	33.0	0.5	198	3	7,643.7	63.8	21.1	-1.4	-6.6
Tibet Aut. Region	Southwest	0.2	0.0	1	0	773.0	62.1	0.2	-0.1	-60.4
Xingjiang Uygur Aut. Region	Northwest	12.7	0.5	76	3	1,978.2	63.4	6.7	-1.9	-28.4
Yunnan Province	Southwest	30.8	0.3	185	2	2,502.0	57.6	6.4	-1.0	-15.5
Zhejiang Province	East	211.5	0.2	1,269	1	9,426.4	62.9	154.7	-0.2	-0.1
Total		-	-	83,534	4,634	217,626.3	-	1,171.4	-72.8	-6.2%

Notes: Monthly information on provincial progression of the pandemic, including new cases, new deaths as well as total cumulative cases and deaths were obtained from Chinese National Health Commission. Monthly province-level statistic on Chinese exports are published by the General Administration of Customs of the People's Republic of China. , Monthly statistics on the number of international cases are sourced from the European Centre for Disease Prevention and Control. Lastly, FVA data to calculate the GVC contagion measure stems from the UNCTAD-Eora Global Value Chain Database.

Covid Economics 53, 23 October 2020: 72-109

In column (6), we present the sum of trade-weighted international case counts that each province faces in the destination countries of its exports.[8] Total Chinese export exposure to trade-weighted international cases amount to nearly 220,000 cases, while provincial level exposure ranges from 773 to 16,000. We interpret this as evidence that each province faces a significant amount of export exposure to foreign infections and is likely to suffer the ramifications of the international demand shock.

In column (7), we report our GVC contagion variable (see Equation (3)), which measures the degree of industry-specific exposure to all international Coronavirus infections. For the purposes of this summary, we determine the provincial-level exposure to GVC contagion by weighting our industry-specific variable with each province's trade composition across these affected industries. Again, we use the first six months of 2019 to determine these trade weights predating the Coronavirus outbreak. The data show that, since every Chinese province is engaged in exports of commodities that are exposed to the disruptions of the underlying international supply chains, each region is similarly vulnerable to GVC contagion in aggregate. Our measure ranges from 46.3 in the Qinghai province to 67.7 in Chongqing.

We combine these intensity measures of the pandemic-induced supply and demand shocks with information on the total and relative year-over-year change in provincial exports between the first halves of 2019 and 2020 (columns (9) and (10)). The data provide preliminary evidence that domestic confirmed Coronavirus cases, trade-weighted international infections, and GVC contagion all negatively correlate with the total change in provincial exports. In other words, greater intensity of the local outbreak and higher levels of exposure to the foreign cases, either directly through the demand channel or indirectly through the disruption of the GVC network, coincide with greater losses in exports during the first half of 2020. Interestingly, Table 1 also shows that provinces with a more severe local outbreak face greater exposure to foreign cases, both directly through demand for its own exports to these countries and indirectly through the 'infected' value added produced by foreign countries participating in the GVC network that provides intermediate inputs to Chinese exporters.

Overall, Table 1 shows that Chinese exports during the first half of 2020 fall by $72.8 billion relative to the previous year. This total reduction in exports amounts to a 6.2% year-over-year loss. As Figure 1 shows, however, this may be a very conservative estimate of the pandemic-induced losses in trade, as Chinese exports grew by over $4 billion per month on average in 2019. Based on this 2019 trajectory, one would have expected a 25% year-over-year increase in Chinese exports by 2020 in the absence of the pandemic.

[8] Trade weights are based on provincial exports during the first half of 2019.

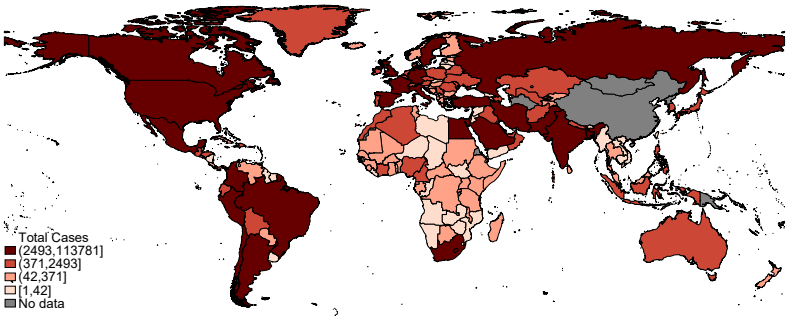
Covid Economics 53, 23 October 2020: 72-109

Shifting our perspective, we plot Figures 2a through 2c to look at the data through the lens of foreign countries importing Chinese products. These maps depict the country-specific total number of confirmed Coronavirus cases during our sample period (Figure 2a), as well as the trade-weighted exposure to Chinese infections through imports from China (Figure 2b) and the relative year-over-year change in these imports (Figure 2c). Figure 2a demonstrates that international cases during the first half of 2020 are largely concentrated in North America, Europe, and South America, while many African nations report smaller numbers of total infections. In contrast, Figure 2b illustrates that global exposure to the Chinese supply disruption (through Chinese exports) is more evenly distributed. South America, the Democratic Republic of Congo, Tunisia, France and Poland, Zambia and several countries neighboring China (i.e. India, Pakistan, Thailand, Laos, and Vietnam, among others) appear most vulnerable to the Chinese supply disruption. North America, Russia, Germany, Italy, and Spain, and many African nations, including Morocco, Algeria, Egypt, Sudan, Angola, and South Africa, among others, also face above average levels of exposure.

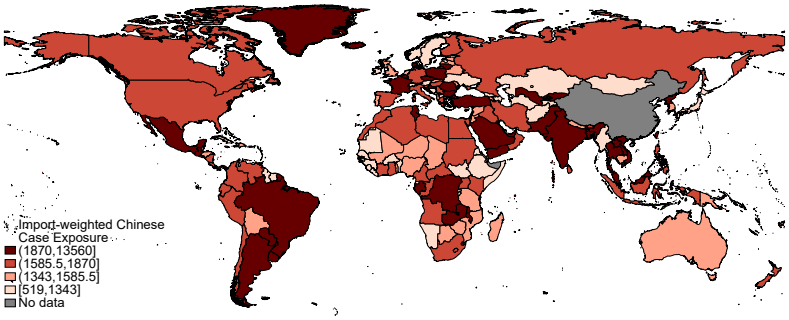
Figure 2c links these pandemic patterns to country-specific changes in imports from China and shows that North and South American countries, as well as nations located in the northern and southern regions of Africa, or those neighboring China experience the largest relative reductions in imports from China. Somewhat unexpectedly, many European nations, which face a large number of domestic infections and/or a high degree of exposure to the Chinese supply disruptions, indicate very moderate reductions in imports from China (or even slight increases).

Next, we consider the dynamic dimension of our data and trace the year-over-year monthly changes in Chinese regional-level exports and continental imports from China over time (see Figures 3a through 3f). The bar charts plot the progression of trade adjustments in January, March, and May of 2020 against the number of new domestic cases as well as the level of exposure to new international infections. The figures show that early pandemic-induced changes in Chinese exports, both from the Chinese and international perspectives, are more correlated with the local number of infections than the exposure to overseas cases. That is, while all Chinese regions face some exposure to new international Coronavirus cases, East and South Central China, for example, experience the highest numbers of new infections and also report the largest reductions in exports (Figures 3a and 3c). Similarly, from the international perspective, all continents appear to face a relatively even level of exposure to Chinese cases (through imports from China), but Asia, North America, and Europe illustrate largest reductions in imports from China while also reporting the highest numbers of new infections in January and March 2020 (Figures 3b and 3d).

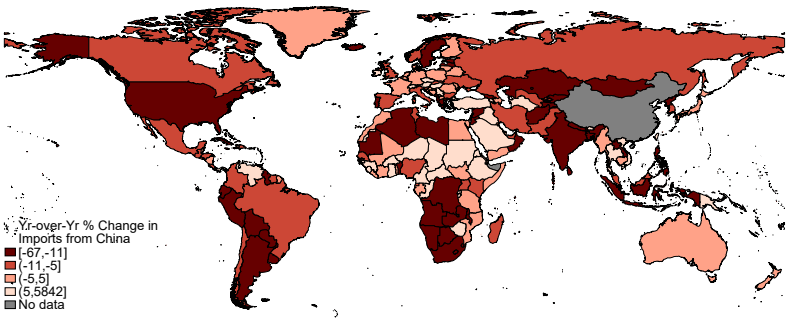
Covid Economics 53, 23 October 2020: 72-109



(a) Total International Covid-19 Cases (Jan.-June 2020)



(b) International Exposure to Chinese Cases (Jan.-June 2020)



(c) International Change in Imports from China (Jan.-June 2019 to 2020)

Figure 2: Int'l Distribution of Covid-19 Cases and Changes in Imports from China

Covid Economics 53, 23 October 2020: 72-109

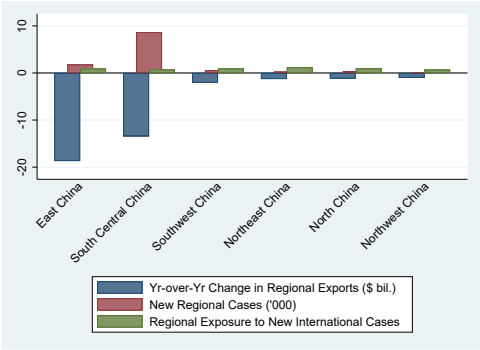
By May 2020, the number of new Chinese infections has fallen drastically and the previously noted correlations are much less apparent. While South Central and East Chinese provinces continue to experience the largest year-over-year reductions in exports, the North and Northeastern provinces report the largest number in new infections in May 2020 (see Figure 3e). Interestingly, European countries, which reported some of the largest reductions in imports from China earlier in 2020, experience a notable year-over-year rise in imports from China by May 2020, despite having a significant number of new infections (Figure 3f).

Lastly, we consider the industry dimension of our data. Of particular interest, of course, is the GVC contagion that each Chinese industry faces as well as the primary contributors to this likely disruptions of international supply chains. To this end, Table 2 lists the top ten most exposed (Panel A) and least exposed (Panel B) industries as well as their respective GVC contagion values. As columns (1) and (2) of Table 2 show our measure of GVC contagion ranges from 32.4 for the least exposed industry (Pulp of wood and recovered paper and paperboard) to 74.5 for the most exposed sector (Aircraft, spacecraft, and parts thereof). While this range suggests a fair amount of dispersion, column (2) of Table 2 also shows that we calculate a GVC contagion value above 60 for 37 of 43 Chinese export industries. That is, almost all industries are expected to face a significant degree of disruption to their international supply chain network.

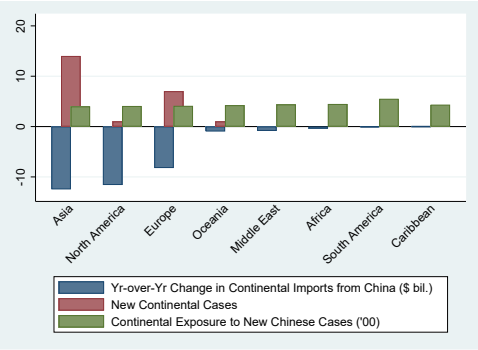
The narrow dispersion of industry-specific GVC contagion comes from the fact that almost all Chinese export industries procure over 80% of their international intermediate inputs from a small set of dominant foreign suppliers. Among these manufacturing hubs that supply intermediate inputs to Chinese exports are the United States, South Korea, Japan, Germany, Italy, Brazil, and to some extent Russia, India, and the United Kingdom (see columns (3) and (4) of Table 2). Because of these similar dependencies, most Chinese export industries face a similar degree of GVC contagion.

As suggested by Baldwin and Freeman (2020), manufacturing sectors appear more vulnerable to GVC contagion than exports of other sectors. We note that there are several key manufacturing sectors among the top ten most exposed industries. These include manufactured aircraft and spacecraft, vehicles, railway, ships, nuclear reactors, machinery, and mechanical appliances, among others. In contrast, as one might expect, the group of least exposed Chinese export industries includes basic commodities, such as ores, slag, and ash, mineral fuels and oils, as well as iron and steel and animal and vegetable products (see column (1) of Table 2). Some industries that do not necessarily fit this pattern include man-made fibres and textiles, fabrics, as well as furniture.

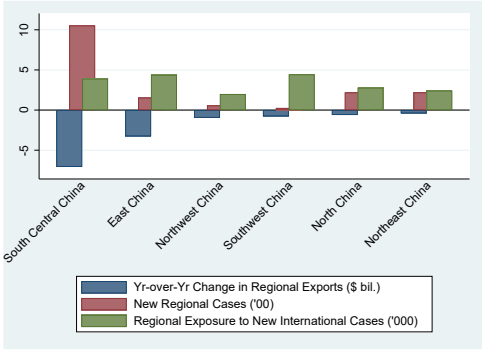
Covid Economics 53, 23 October 2020: 72-109



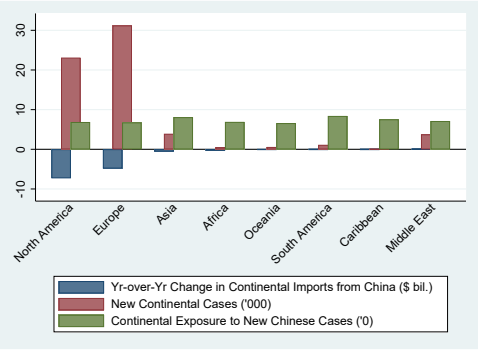
(a) January



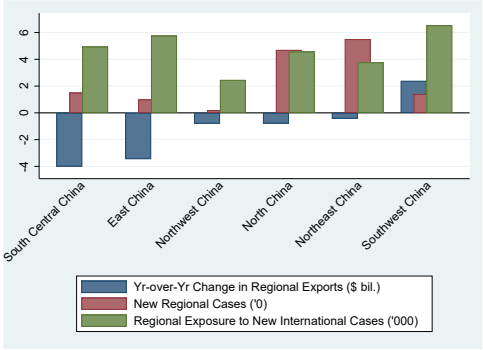
(b) January



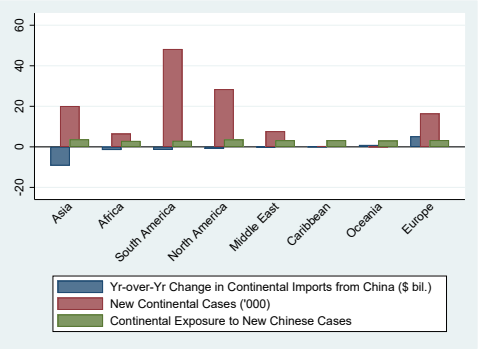
(c) March



(d) March



(e) May



(f) May

Figure 3: Monthly Exposure to Chinese and Int'l Covid-19 Cases and Changes in Trade

Covid Economics 53, 23 October 2020: 72-109

Table 2: GVC Contagion by Chinese Export Industry

(1) Industry	(2) GVC-C	(3) Top 5 Contributing Foreign Countries	(4) GVC-C Share of Top 5 (%)	(5) Y-o-Y Change in Exports (Jan.-June, 2019-20) ($ bil.)	(6) (%)
Panel A: Top Ten					
Aircraft, spacecraft, and parts thereof	74.5	U.S., South Korea, Japan, Germany, Italy	84.4	-0.3	-15.2
Raw hides, Articles of leather, Furskins, Prepared feathers, etc.	74.0	South Korea, U.S., Italy, Japan, Germany	87.1	-5.4	-27.8
Vehicles other than railway and parts and accessories thereof	70.2	U.S., South Korea, Japan, Germany, Brazil	84.5	-4.5	-12.6
Nuclear reactors, machinery and mechanical appliances and parts thereof	69.7	U.S., South Korea, Japan, Germany, Brazil	83.2	-9.6	-4.7
Ships, boats and floating structures	67.5	U.S., South Korea, Japan, Germany, Italy	84.0	-1.8	-15.9
Base metals (other than iron and steel), and articles thereof, etc.	66.7	South Korea U.S., Brazil, Japan, Germany	82.6	-2.5	-13.6
Tools, implements, cutlery, other parts and articles of base metal	66.6	U.S., South Korea, Japan, Germany, Brazil	82.4	-1.9	-11.3
Railway or tramway locomotives, track fixtures and fittings, etc.	66.5	U.S., South Korea, Japan, Germany, Brazil	82.7	-1.0	-21.2
Tobacco and manufactured tobacco substitutes	66.4	U.S., South Korea, Japan, Germany, Italy	84.2	-0.2	-36.6
Printed books, newspapers, pictures, etc.	65.8	U.S., South Korea, Japan, Germany, Italy	84.3	-0.4	-23.1
Panel B: Bottom Ten					
Iron and steel and articles thereof	61.7	U.S., South Korea, Brazil, Japan, India	81.0	-6.2	-11.5
Animal or vegetable fats and oils, etc.	61.1	U.S., South Korea, Japan, Germany, Russia	79.8	0.1	27.0
Residues and waste from the food industries, animal fodder	60.4	U.S., South Korea, Japan, Russia, Germany	81.1	0.1	3.8
Man-made staple fibres, filaments and textile materials, etc.	60.4	U.S., South Korea, Japan, Italy, Germany	81.8	-7.0	-24.4
Knitted or crocheted fabrics	59.8	U.S., South Korea, Japan, Italy, Germany	82.2	-2.3	-26.1
Preparations of meat, of fish or of crustaceans, etc.	59.8	U.S., South Korea, Japan, Russia, Germany	81.4	-0.5	-11.8
Furniture; bedding, mattresses, lamps and lighting fittings, etc.	59.7	U.S., South Korea, Japan, Russia, Germany	81.0	-5.8	-10.7
Ores, slag and ash	58.6	U.S., South Korea, Japan, Germany, Italy	79.6	0.2	25.4
Mineral fuels, mineral oils and products thereof, etc.	49.8	U.S., South Korea, Japan, Iran, Germany	78.3	-4.0	-17.5
Pulp of wood, recovered (waste and scrap) paper or paperboard, etc.	32.4	U.S., South Korea, UK, Brazil, Japan	73.3	-0.0	-12.5

Notes: Our calculation of GVC contagion (see Equation (3)) depends on the data on foreign-country-to-Chinese-industry Foreign Value Added provided by the UNCTAD-Eora Global Value Chain Database (Casella et al., 2019). We merge the FVA data with Chinese trade statistics based on a two-digit HS classification. Our concordance between the Eora GVC database and the GACC trade data aggregates industries across the two datasets and yields 43 matching export industries. The concordance is available upon request.

A calculation of the year-over-year changes (from January through June 2020 relative to 2019) in Chinese exports across the ten most and least exposed industries reveals that more exposure to GVC contagion positively correlates with the losses in exports, both in terms of total and relative changes. Column (5) shows that year-over-year changes in exports for the ten most exposed industries ranges from -$0.2 billion to -$9.6 billion, whereas the ten least exposed industries experience export changes ranging from -$7.0 billion to $0.2 billion. In relative terms, export losses for the ten most exposed industries range from 4.7% to 36.6%, whereas the least exposed industries report relative year-over-year changes in exports ranging from -26.1% to 27.0% (see column (6) of Table 2).

5 Results

Building on the patterns observed in the raw data, our empirical analysis consists of three components. First, we conduct our baseline analysis following Equation (2) and estimate the elasticity of Chinese exports with respect to new domestic and international Covid-19 infections reported in the province of origin and country of destination. We test the sensitivity of our estimates against model alterations, including alternative pandemic measures and fixed effects specifications, as well as more relaxed sample restrictions. Our baseline findings point to significant and robust disruptions in Chinese exports in response to the local and international outbreak of the novel Coronavirus.

Next, we explore the heterogeneity of these baseline pandemic effects on Chinese exports across various country and commodity characteristics. New infections in OECD and other high income countries seem to cause smaller disruptions of imports from China than in other lower income countries. Geography matters as well. As suggested by Figure 2c, Chinese exports to neighboring countries experience greater disruptions from international infections.

Lastly, we estimate Equation (4) in an effort to differentiate the international demand effect from the disruptions resulting from the pandemic-induced GVC contagion. The dynamic coefficient estimates suggest that the Coronavirus pandemic impacts trade through all three of the discussed transmission channels; and does so in a staggered fashion over the first 3 month after a rise in new infections.

Across all specifications we correct our standard errors for two-way clusters controlling for correlations at the bilateral province-to-foreign-country and commodity levels. Depending on the estimation sample, this correction adjusts standard errors across 1900 to 5000 bilateral trade province-country pairs and 43 to

Covid Economics 53, 23 October 2020: 72-109

97 commodity clusters. Our inference is robust to alternative standard error specifications.

5.1 Baseline Estimates

We begin by estimating Equation (2). Panel A of Table 3 shows the baseline coefficient estimates. We build from a set of parsimonious specifications that only consider a one-sided pandemic effect (columns (1) and (2)) or two-sided impact with a limited set of fixed effects (column (3)) to the preferred baseline specification given in column (4). This specification produces the elasticities of Chinese exports with respect to local and foreign Coronavirus cases while controlling for bilateral province-to-foreign-country unobservable characteristics, time-invariant differences across industries, and macroeconomic trends. Across all of the parsimonious and preferred specifications the coefficients carry the expected negative sign and are statistically significant at the 1% level. Our estimates suggest that a 1% rise in new Chinese provincial Coronavirus cases reduces exports from that province by 3.7%. Foreign cases appear even more detrimental to trade. A 1% rise in new foreign cases in the destination country lowers it's imports from China by 4.6% (see, for example, column (4) in Panel A of Table 3). These findings are consistent across alternative fixed effects specifications (column (3)) and considering only one of these shocks (columns (1) and (2)).

Reassuringly, these pandemic effects are also robust to alternative measures of the severity of the Coronavirus outbreak. Column (5) of Panel A of Table 3 shows that a 1% increase in new Covid-19 related deaths in Chinese provinces reduces exports by 12.5%, whereas 1% rise in new pandemic-related deaths in the destination country lowers imports from China by 4.6% - equivalent to the new case effect. Column (6) demonstrates the impacts of cumulative Coronavirus cases. Both the local and foreign accumulation of Coronavirus infections reduces Chinese exports. A 1% rise in provincial or foreign country cumulative cases drops Chinese exports by 8.1% and 3.5%, respectively.

These results are not driven by our sample restrictions excluding the U.S., due to the ongoing trade war, or other foreign countries, for which we only observe a few months of case data. While including the latter set of countries reduces the magnitude of the foreign case effect, the coefficients remain statistically significant at the 1% level. For this sample, a 1% increase in local and foreign cases reduces Chinese exports by 4.3% and 2.5% respectively (see column (7) of Panel A of Table 3). Even the inclusion Chinese exports to the United States does not overturn our baseline results. As the coefficients reported in column (8) of Panel A show, local and foreign outbreaks of the novel Coronavirus have economically and statistically

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

Table 3: Pandemic Effects on Chinese Exports

Dependent Variable: Inverse Hyperbolic Sine of Exports	(1) Restrict. Sample	(2) Restrict. Sample	(3) Restrict. Sample	(4) Restrict. Sample	(5) Restrict. Sample	(6) Restrict. Sample	(7) Exclude U.S. only	(8) Incl. U.S.
Panel A: Baseline Estimates								
New Chinese Cases	-0.037		-0.037	-0.037			-0.043	-0.042
	(0.000)		(0.000)	(0.000)			(0.000)	(0.000)
New Int'l Cases		-0.046	-0.046	-0.046			-0.025	-0.024
		(0.000)	(0.000)	(0.000)			(0.000)	(0.000)
New Chinese Deaths					-0.125			
					(0.000)			
New Int'l Deaths					-0.046			
					(0.000)			
Cumulative Chinese Cases						-0.081		
						(0.000)		
Cumulative Int'l Cases						-0.035		
						(0.001)		
N	1683468	1683468	1683468	1683468	1683468	1683468	3675150	3717810
R^2	0.438	0.438	0.438	0.457	0.457	0.457	0.445	0.449
Province FE	Y	Y	Y	N	N	N	N	N
Country FE	Y	Y	Y	N	N	N	N	N
Bilateral FE	N	N	N	Y	Y	Y	Y	Y
Commodity FE	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Heterogeneity Analysis

	Country Type		Geography		Commodity Type			
	OECD	High Income	Contiguous Neighbor	Inverse Distance	Pharma-ceuticals	Air Borne	DVX	FVA
New Chinese Cases	-0.036	-0.036	-0.039	-0.043	-0.038	-0.037	-0.028	-0.026
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.016)	(0.017)
x Z_j	-0.004	-0.002	0.015	25.499	0.049	-0.013	0.012	0.005
	(0.545)	(0.762)	(0.130)	(0.343)	(0.000)	(0.698)	(0.181)	(0.058)
New Int'l Cases	-0.061	-0.063	-0.033	-0.026	-0.048	-0.044	-0.044	-0.042
	(0.000)	(0.000)	(0.000)	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)
x Z_j	0.040	0.035	-0.058	-104.071	0.122	-0.050	0.005	0.002
	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.002)	0.526	0.526
R^2	0.457	0.457	0.457	0.457	0.457	0.457	0.526	0.526
N	1683468	1683468	1683468	1683468	1683468	1683468	875304	875304
Bilateral FE	Y	Y	Y	Y	Y	Y	Y	Y
Commodity FE	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: P-values are reported in the parenthesis. The underlying standard errors are two-way clustered controlling for correlations at the bilateral province-to-foreign-country and commodity levels. In panel A, the regression results presented in columns (1) through (3) include separate province and foreign country fixed effects, whereas the regression underlying the results given in columns (4) through (8) include bilateral province-foreign-country fixed effects. All specifications include commodity and time fixed effects. In panel B, we investigate the heterogeneity of the pandemic effects on Chinese exports across subgroups of countries (i.e. OECD, high income) with respect to geography (i.e. countries contiguous to China or interacted with inverse distance from China) and commodities (i.e. pharmaceutical products or products primarily transported by air). Across columns (1)-(6) of Panel A and all columns of Panel B, we restrict the sample excluding exports to the U.S. and countries for which we observe less than six month of case data on the Covid-19 outbreak. This restricted sample consists of a total of 1927 bilateral Chinese province to foreign country pairs, for which we observe a positive value of exports for at least one of 97 commodities over at least one of 18 months from January 2019 to June 2020. Columns (7) and (8) of Panel A include other foreign countries with fewer Coronavirus case observations and the United States, respectively.

significant ramifications for international trade.

5.2 Heterogeneity Analysis

Given the wide set of political responses to the pandemic and varying capacities to deal with an outbreak as well as varying dependencies on international trade, one should not expect these baseline estimates to be representative of the pandemic impact on imports of every commodity or by every country. To shed light on these potential idiosyncrasies, we conduct a heterogeneity analysis of the local and foreign case effects on Chinese exports. Specifically, we explore the variation in pandemic impacts on Chinese exports along the foreign country and commodity dimensions. Expanding the baseline specification (Equation (2)) with two simple interaction terms ($ci_{pt}Z_l$ and $ci_{jt}Z_l$ where $l = j$ or k), we delineate the effects on new infections across different economic or geographical attributes of the destination countries (Z_j) as well as a few select commodity characteristics (Z_k). The results are presented in Panel B of Table 3.

We begin by differentiating the new case effects on Chinese exports to OECD and other high-income countries from those experienced by other, lower income countries. Columns (1) and (2) of Panel B of Table 3 show that a rise in new Chinese infections has essentially the same impact on exports irrespective of the destination country's level of income. One way to interpret this finding is that in the face of local supply disruptions Chinese exporters do not discriminate across foreign buyers from high to low income countries. In contrast, we find that the import effects of new infections in the destination countries depends on the country's level of development. Our coefficients suggest that the impact on imports from China by non-OECD and low to medium income countries is more than twice as large as the trade disruption experienced by OECD and other high income countries. A 1% rise in new infections in an OECD country, for example, reduces imports from China by 2.1%(=-0.061+0.04), whereas the same relative increase in new case in a non-OECD country reduces imports from China by 6.1% (see column (1), Panel B of Table 3). This indicates that the pandemic-induced demand shock for traded products is much more severe in low-income countries.

Geography also plays a role in the determination of the pandemic-induced trade effects. Similar to the influence of income levels, however, geography only shapes the import effects of new infections in the destination country. While Chinese provincial infections have similar effects on exports to neighboring countries and those at greater distances, the impact of destination country cases on imports from China decreases with distance to its trade partner (see columns (3) and (4), Panel B of Table 3). More specifically,

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

we find that a 1% increase in new Coronavirus infections in a country neighboring China, such as India, Pakistan, or Vietnam, causes Chinese exports to these countries to drop by close to 9%(=-0.033+(-0.058)) on average. This effect is nearly three times as large as the estimated coefficient for non-neighboring countries.

With respect to commodity characteristics, we start by investigating whether Chinese exports of pharmaceuticals are affected differently from other types of traded products. As expected, the generally negative impact of the the pandemic on Chinese exports is completely overturned for trade in pharmaceuticals and other medical equipment. While the effects of new local Chinese infections is essentially neutral on exports of pharmaceuticals, a rise of new infections in the destination country stimulates imports in medical goods from China. Specifically, Chinese exports of pharmaceuticals rise by 1.1% and 7.4% for every 1% increase new Chinese and new foreign cases, while the average Chinese export falls by 3.8% and 4.8%, respectively (see column (5), Panel B of Table 3).

Aside from isolating the effects on medical goods, we also consider the influence transportation on shaping the Covid-19 pandemic effect on Chinese exports. Given the significant restrictions on international travel of persons, which has lead to a 70% decline in air traffic that also facilitates traded products (World Trade Organization, 2020a), we estimate whether this policy response has the unintended consequence of significantly reducing Chinese exports of commodities that are typically transported by air. As one might expect, we find that the elasticity of exports of likely airborne products, such as jewelry or life animals, with respect to new infections in the destination country is more than twice as large as the elasticity of exports in other, likely seaborne commodities (see column (6), Panel B of Table 3).

Lastly, we conclude this first component of our heterogeneity analysis by investigating whether individual linkages in the GVC network influence the impact of new Chinese and foreign infections. This analysis is a precursor to the primary analysis that aims to capture the full GVC contagion effect. Specifically, we test whether Chinese exports are more or less sensitive to new infections 1) if these Chinese exports provide a higher value added to the exports of the foreign destination country (column (7), Panel B of Table 3) or 2) if these exports are flowing to a foreign country that in turn provides a higher value added to Chinese production through its own exports to China (column (8), Panel B of Table 3). While all of the coefficients on the relevant interactions terms are positive, only one is statistically significant at the 10% level. This statistically significant coefficient suggests that new local infections in China have a marginally smaller effect on exports to foreign countries that, in turn, provide more FVA to Chinese exporters. While this could be interpreted as strategic behavior that favors trade partners that in turn provide important inputs into

Chinese production, the evidence is suggestive at best. Overall, we take the results presented in columns (7) and (8) of Panel B of Table 3 to mean that individual links in the GVC networks play only a minor role in shaping the pandemic-induced trade outcomes. We look to our primary results to shed more light on the cumulative GVC contagion effect.

5.3 Primary Estimates of Domestic Supply, Int'l Demand, and GVC Contagion Effects

Our final and primary analysis investigates the individual transmission mechanisms underlying the previously established pandemic effects on Chinese exports. Estimating Equation (4), we distinguish between the domestic supply effect (see Panel A of Table 4), the international demand effect (see Panel B of Table 4), and the GVC contagion effect (see Panel C of Table 4) on Chinese exports as well as their respective timing. As previously mentioned, the merger between the FVA data and our statistics on Chinese exports and global infections to calculate the GVC contagion measure limits our estimation sample to 43 matching industries and 1900 bilateral Chinese-province-foreign-country pairs reducing the number observations from 1.68 million to around 875,000. Because of this sample restriction, we begin by re-estimating our baseline specification. The coefficient estimates in column (1) of Table 4 are statistically significant at the 5% level or below and show that new Chinese and international infections continue to reduce Chinese exports.

Interestingly, the point estimates remain virtually unchanged when we include our contemporaneous GVC contagion measure (see column (2)). That is, during the month that a Chinese province experiences a 1% rise in new Coronavirus infections the ensuing domestic supply shock reduces Chinese exports by 2.6%. Similarly, the international demand shock due to a rise in destination country cases is immediate and large. A 1% rise in new infections reduces the demand for imports from China by 4.2%. In contrast, the potential GVC contagion effect from the disruption of foreign production does not take hold during the month of a rise in international infections. The point estimate reported in column (2) of Table 4 is positive and statistically insignificant.

The results reported in columns (3) through (5) of Table 4 consider the dynamic effects of these three shocks from the first to the third month after a rise in Chinese and international infections. As one might expect, the disruption of Chinese export supply worsens during the first month after a rise in new infections and then slowly dampens over the next two month. A one standard deviation increase in new Chinese infections (see Panel D of Table 5), for example, reduces Chinese exports by more than 11%(=-0.044*2.595)

Covid Economics 53, 23 October 2020: 72-109

Covid Economics 53, 23 October 2020: 72-109

Table 4: Supply and Demand Effects of the Covid-19 Pandemic on Chinese Exports

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Individual Effects				All Effects			
	Baseline	Con-tempo-raneous	1^{st} Lag	2^{nd} Lag	3^{rd} Lag	IHS of Exports	IHS of Exports	Exports	Exports
Panel A: Domestic Supply Effects									
New Chinese Cases$_{pt}$	-0.026	-0.026				-0.011	-0.020	-61,136	-35,817
	(0.020)	(0.028)				(0.263)	(0.004)	(0.021)	(0.027)
New Chinese Cases$_{pt-1}$			-0.044			-0.037	-0.055	-91,506	-56,967
			(0.002)			(0.000)	(0.000)	(0.011)	(0.008)
New Chinese Cases$_{pt-2}$				-0.026		-0.012	-0.014	35,951	17,518
				(0.075)		(0.375)	(0.142)	(0.117)	(0.192)
New Chinese Case$_{pt-3}$					-0.001	0.021	0.028	65,531	51,164
					(0.966)	(0.163)	(0.015)	(0.151)	(0.087)
Panel B: International Demand Effects									
New Int'l Cases$_{jt}$	-0.042	-0.042				-0.042	-0.011	-108,624	-92,018
	(0.000)	(0.000)				(0.000)	(0.101)	(0.003)	(0.005)
New Int'l Cases$_{jt-1}$			-0.029			-0.019	-0.026	-9,521	20,681
			(0.002)			(0.020)	(0.000)	(0.703)	(0.278)
New Int'l Cases$_{jt-2}$				0.016		0.037	0.017	81,663	72,851
				(0.097)		(0.000)	(0.008)	(0.030)	(0.039)
New Int'l Cases$_{jt-3}$					0.037	0.029	0.014	83,029	92,118
					(0.002)	(0.012)	(0.129)	(0.099)	(0.050)
Panel C: GVC Contagion Effect									
GVC-C$_{kt}$		0.004				0.007	0.010	-44,236	-26,270
		(0.705)				(0.462)	(0.283)	(0.287)	(0.286)
GVC-C$_{kt-1}$			-0.054			-0.055	-0.056	-70,941	-49,081
			(0.001)			(0.000)	(0.000)	(0.065)	(0.058)
GVC-C$_{kt-2}$				-0.048		-0.035	-0.033	-23,357	-25,997
				(0.001)		(0.010)	(0.003)	(0.252)	(0.087)
GVC-C$_{kt-3}$					-0.018	-0.024	-0.026	-4,643	-9,505
					(0.333)	(0.223)	(0.136)	(0.874)	(0.585)
N	875304	875304	875304	875304	875304	875304	2061972	875304	2061972
R^2	0.526	0.526	0.526	0.526	0.526	0.526	0.519	0.157	0.139
Bilateral FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Commodity FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel D: Summary Statistics

Restricted Sample	Mean	SD	IQR	5th Perc.	95th Perc.
New Chinese Cases$_{pt}$	3.419	2.595	4.836	0	7.137
New Int'l Cases$_{jt}$	3.837	3.154	6.534	0	8.712
GVC-C$_{kt}$	10.681	3.427	3.883	6.483	17.219
Unrestricted Sample					
New Chinese Cases$_{pt}$	3.469	2.612	4.836	0	7.137
New Int'l Cases$_{jt}$	2.828	3.065	5.497	0	8.336
GVC-C$_{kt}$	10.698	3.462	3.883	6.483	17.219

Notes: P-values are reported in the parenthesis. The underlying standard errors are two-way clustered controlling for correlations at the bilateral province-to-foreign-country and commodity levels. The restricted estimation sample consists of 43 traded commodities observed over 18 months for 1,900 bilateral Chinese province-to-foreign-country pairs (columns (1)-(6) and (8)) and 5,119 bilateral pairs for the unrestricted sample, including the U.S. and other countries for which we observe Covid-19 data only after January 2020 (columns (7) and (9)). Column (1) presents the baseline estimates for this new sample. Starting with Column (2), we also control for the GVC contagion effect. Columns (2) through (5) report coefficient estimates for each of the separately estimated dynamic effects ranging from the contemporaneous impacts to the impact after three month. Columns (6) through (10) report coefficient estimates for each of the jointly estimated dynamic effects. For columns (1) through (8), the dependent variable is given by the IHS of Chinese exports, whereas for columns (9) and (10) the dependent variable is given by total Chinese exports in U.S. $.

one month after the worsening of the Coronavirus outbreak, around 7%(=-0.026*2.595) two month later, and less than 0.5%(=0.001*2.595) three month after the initial domestic supply disruption.

The effects of a shock to international demand are much less persistent and in fact reverse over time. While the immediate impact of a one standard deviation increase in new destination country infections on imports from China is large at -13.2%(=-0.042*3.154), it falls to -9.1%(=-0.029*3.154) one month later. By the second month, the initially negative demand shock has worn off and imports from China actually rise with a greater number of past destination country infections. This perhaps is the result of the supply disruption in foreign countries that consequently become more dependent on and demand more imports from China. Specifically, we find that two to three month after a one standard deviation increase in new destination country cases imports from China increase by more than 5%(=0.016*3.154) to 11%(=0.037*3.154), respectively (see columns (4) and (5), Panel B of Table 4).

The effects of GVC contagion arise one month after the initial increase in new foreign cases and are relatively persistent during the second and third months thereafter. Based on our estimates, we find that a one standard deviation increase in our GVC contagion measure (see Panel D of Table 5) reduces Chinese exports by more 18.5%(=-0.054*3.427) one month after the initial outbreak and 16.4%(=-0.048*3.427) the following month. The coefficient on the third lag suggests a reduction of 6.2%(=-0.018*3.427), but is statistically indistinguishable from zero (see columns (3) through (5), Panel C of Table 4).

Next, we consider the joint estimation of these dynamic pandemic effects on trade (see column (6)). The domestic supply shock in China continues to exert a negative influence on Chinese exports during the month of a rise in new Coronavirus infections and two months thereafter. However, only the coefficient on the first month lag is statistically significant at the 1% level. The international demand shock on Chinese exports is economically and statistically significant across the immediate and lagged effects. As before, the pandemic shock in the destination countries reduces imports from China during the month of the outbreak and one month thereafter, but reverses during the second and third months after the initial rise in foreign cases. It appears that foreign countries become more reliant on imports from China as their domestic supply is being disrupted by the pandemic. Similar to the individual analyses, the joint estimation shows that the GVC contagion is not immediate but takes hold one to two months after the outbreak in foreign countries.

We test the sensitivity of these results against our initial sample restrictions as well as our inverse hyperbolic sine transformation of Chinese exports. Including the U.S. and countries for which we observe new Coronavirus cases only after January, 2020 largely reiterates and strengthens our primary findings.

Covid Economics 53, 23 October 2020: 72-109

The domestic supply shock in China once again becomes statistically significant during the month of a rise in local cases, but also starts to reverse after a three-month lag. The international demand shock and GVC contagion coefficients continue to exhibit the aforementioned patterns (see column (7) of Table 4).

The regression analysis of the triple pandemic effect on the untransformed value of Chinese exports also produces similar estimate patterns with and without the aforementioned sample restrictions (see columns (8) and (9) of Table 4). The interpretation of our coefficient estimates, however, changes. Considering the restricted sample analysis (column (8) of Table 4), we find that, on average, a one standard deviation increase in new Chinese Coronavirus infections reduces Chinese exports of a given commodity from a given province to a given foreign country by around $150,000 during the month of the outbreak and nearly $240,000 one month later. Thereafter, the domestic supply shock evaporates quickly and becomes statistically insignificant. On the demand side, a one standard deviation increase in destination country cases reduces the average province-to-foreign-country trade flow of a two-digit commodity by close to $350,000 during the month of the outbreak. The first lagged effect becomes statistically insignificant. Thereafter, a previous rise in foreign infections stimulates the demand for imports from China. On average, a one standard deviation increase in destination country cases raise the demand for Chinese exports by around $250,000 two to three month after the outbreak. The estimated dynamic GVC contagion effects are all negative but only statistically significant during the first month after the rise in global infections. This point estimate suggests that a one standard deviation increase in the 'infection' of the supply chain network reduces Chinese-province-to-foreign country exports of a two-digit commodity by nearly $250,000 on average. Expanding the sample to include the U.S. and other countries with limited pandemic information yields very similar qualitative and quantitative evidence of the triple effect of the pandemic (see column (9) of Table 4).

6 Discussion and Limitations

Overall, these results suggest that all three channels contribute to the pandemic-induced trade effects, but also indicate that the timing and magnitudes of the domestic supply shock, international demand shock and impact of GVC contagion vary distinctly. In Table 5, we determine the relative trade effects implied by the data and our regression estimates reported in columns (8) and (9) of Table 4. We calculate the predicted relative changes in Chinese exports and break out the monthly dynamics starting from the immediate im-

Covid Economics 53, 23 October 2020: 72-109

pact through the predicted changes three months after a rise in new infections. We use these dynamic counterfactual estimates to produce a breakdown of the individual relative contributions across the three supply and demand shocks for each of the immediate and lagged time periods. Statistics reported in Panel A of Table 5 are based on the restricted sample, whereas Panel B presents this information based on the unrestricted sample.

Column (1) of Table 5 shows the predicted percentage changes in Chinese exports combining the effects of the domestic supply shock, international demand shock and GVC contagion. In aggregate, our estimates suggest that the pandemic has reduced Chinese exports by as much as 45%. For several reasons, however, we interpret this estimate as an upper bound to the potential pandemic-induced losses in Chinese exports. First, we note that column (1) of Table 5 demonstrates that the vast majority of these losses are incurred immediately or one month after a rise of new infections. As time goes on, we find that Chinese exports tend to partially recover some of those losses. Three month after the initial rise in new global infections, for example, Chinese exports are predicted to rise by more than 5% above the Covid-free counterfactual implying that the initial outbreak stimulates Chinese exports to affected countries after two to three months. Given that our sample ends in June 2020, this gives hope that Chinese exports may recover further by the end of 2020 and total annual losses in exports will likely fall below our benchmark estimate.

Secondly, we note that many of our observations of exports are zero-valued imposing a potentially important limitation for our model. Untructated predictions based on linear regression estimates might suggest negative trade values and consequently overestimate the predicted losses in exports. To test whether this issue plays a significant role, we restrict the sample to bilateral trade flows of two-digit commodities that post a positive export value for every month in our sample and re-estimate our primary model. Reassuringly, many commodities are consistently exported, such that this restricted sample accounts for about 85% of total Chinese exports we observe. The coefficient estimates are qualitatively and quantitatively consistent with our primary results and the predicted pandemic-induced loss in Chinese exports remains relatively stable at about 40%.

Lastly, we note that many of the previously-cited estimates of global losses in exports are considerably lower than our 40% to 45% benchmark, but typically based on year-over-year comparisons in trade. A recent press release by the WTO, for example, predicts global year-over-year annual losses in trade at around 18.5%, up from a pessimistic initial estimate of 32% (World Trade Organization, 2020b). Our data suggest a 17% year-over-year decline in Chinese exports in January and February and an overall year-over-year

Covid Economics 53, 23 October 2020: 72-109

Table 5: Aggregate Pandemic Effect & Individual Shock Contributions

Timing of Impact	(1) Predicted Δ in Exports (%)	(2) Rel. Impact Contributions (%) Supply Shock	(3) Demand Shock	(4) GVC Contagion
Panel A: Restricted Sample (excl. U.S.)				
Immediate only	-27.3	19.0	38.0	43.0
1^{st} Lag only	-24.3	29.5	2.8	67.7
2^{nd} Lag only	2.1	129.0	202.2	-231.2
3^{rd} Lag only	5.7	72.1	40.9	-13.0
Aggregate Impact	**-43.8**	**12.8**	**10.4**	**76.8**
Panel B: Unrestricted Sample (incl. U.S.)				
Immediate only	-25.7	18.7	39.1	42.2
1^{st} Lag only	-23.1	30.5	-7.2	76.8
2^{nd} Lag only	-2.3	-87.4	-166.0	353.4
3^{rd} Lag only	5.2	97.4	48.1	-45.4
Aggregate Impact	**-45.9**	**10.3**	**4.4**	**85.4**

Notes: Predicted levels of exports are based on regressions reported in columns (8) and (9) of Table 4. Column (1) reports the predicted relative changes in Chinese exports against a Covid-free counterfactual. Columns (2) through (4) report the relative contributions of the domestic supply shock, foreign demand shock, and GVC contagion to the predicted change in exports.

reduction of around 5% to 10% during the first half of 2020 depending on the sample. While these statistics are much lower than our benchmark estimate, it is important to note that such year-over-year comparisons ignore the annual growth in exports. In our sample, Chinese exports rose, on average, by about 2.1% per month in 2019 or 25% annually (see Figure 1). Accordingly, an improved estimate of total losses should combine year-over-year losses with the anticipated export growth forgone due to the pandemic. Based on our data, this suggests a rough estimate of Chinese export reductions of about 30% to 35%, which is line with our benchmark estimates, particularly when excluding zero-valued observations.

An interesting question that remains is which of the three pandemic-induced supply and demand shocks drives the aggregate results. To answer this question we decompose our aggregate estimates into the individual contributions across the three transmission channels. Our estimates show that GVC contagion exerts the largest and most consistent negative influence on the change in Chinese exports. In aggregate, the impact of GVC contagion explains 76.8% to 85.4% of the total reduction in Chinese exports depending on the estimation sample. The domestic supply shock in China accounts for around 10% to 13%, while the international demand shock only explains 4.4% in the unrestricted sample and around 10% in the restricted sample. As such, the mechanics underlying the current pandemic effects on trade appear to be very distinct

Covid Economics 53, 23 October 2020: 72-109

from those explaining the Great Trade Collapse in 2008-09. Many researchers have argued that the 2008-09 GTC is largely attributable to the collapse in global demand, particularly that for durable goods, due the financial crisis. In contrast, our estimates suggest that the majority of the pandemic-induced GTC of 2020 can be explained by the unparalleled disruption of GVCs and the reversal of the international demand shock two months after the initial outbreak.

It is important to note that these aggregate statistics mask the importance of the domestic supply and international demand shocks during any of the dynamic period under consideration. While GVC contagion exerts a persistent negative influence on Chinese exports the effects of the shocks to domestic supply and in particular international demand reverse after two to three month. As a result, the domestic supply and international demand shocks offset some of their initially negative impacts and appear as less important contributors to the losses in Chinese exports in aggregate. If we, however, isolate the contemporaneous shock contributions during the month of a rise in new infections, for example, we find a much more even distribution. While the domestic supply shock in China explains around 20% of the immediate reduction in Chinese exports, the international demand shock and ensuing GVC contagion account for around 40% each. One month after the initial rise in infections, the contributions to export losses of the domestic supply shock in China and GVC contagion rises to around 30% and 70%, respectively, while demand shock essentially evaporates. Thereafter, our estimates indicate that the initially negative Chinese supply and international demand shocks are overturned and in fact drive an increase in Chinese exports, while GVC contagion continues to slow export growth. Because of this offsetting effect of the domestic supply and international demand shocks against the persistent negative impact of GVC contagion, the vast majority of the overall reduction in Chinese exports is attributable to the global supply chain disruptions and sets this pandemic apart from other previous impediments to trade.

A final caveat to our results are the facts that the underlying estimation sample only considers Chinese exports and consists 43 industries producing tradable commodities that match the UNCTAD Eora FVA data. Aside from a few sectors producing tradable raw materials and basic commodities, the majority of these industries can be classified as manufacturing. Manufacturing industries, of course, are inherently more susceptible to the disruption of global supply chain networks. Moreover, China has positioned itself as the global manufacturing hub and is very much at the center of many GVCs. Consequently, Chinese exporters provide critical intermediate inputs to foreign producers and are heavily dependent on foreign suppliers. Keeping these facts in mind, it is not surprising that the estimated contribution of GVC contagion

Covid Economics 53, 23 October 2020: 72-109

to the reduction in Chinese exports outweighs the effects of the domestic supply and international demand shocks. If our analysis focused on exports of less globally integrated countries and/or industries with less complex production processes, we would expect the GVC contagion effect to be less dominant.

7 Conclusions

In this study, we have investigated the impact of the 2020 novel Coronavirus pandemic on Chinese exports during the first half of 2020. Our estimates show that Chinese exports are highly sensitive to the domestic and global rise in infections and that the pandemic has imposed a major barrier to international trade. Average Chinese export elasticities with respect to domestic and foreign destination country infections range from -2.5 to -4.6. In aggregate, the pandemic-induced losses in Chinese exports during the first half of 2020 are immediate and may be as large 40% to 45%, but will likely decline by the end of the year.

Across the three potential transmission mechanisms, we find that GVC contagion is the primary driver of these losses in Chinese exports. While the domestic supply and international demand shocks exert a significant negative influence on Chinese exports during the first two months of an outbreak, the impact of GVC contagion persists over three months after the initial rise in Coronavirus infections. Moreover, we find that the pandemic has the unique feature that the international demand shock reverses after about two months after the initial outbreak and stimulates Chinese exports to affected destination countries, perhaps to mitigate their own supply disruptions. The combination of this unique demand shock pattern and dominant GVC contagion effect set the pandemic apart from previous trade shocks of this magnitude, including the primarily demand-driven GTC of 2008-09.

While our study provides some useful insights into the trade effects of the novel Coronavirus pandemic, many important questions remain. Future research may investigate whether the pandemic effects highlighted here are unique to Chinese exports or can be generalized to other countries and other tradable sectors. Moreover, our dynamic short-run analysis illustrates some interesting supply and demand shock patterns that deserve more attention in the medium to long-run. As more data become available, it will be important to understand whether the dominant GVC contagion effect found in our study is short-lived or whether this intense disruption of global supply chain networks has lasting consequences for international trade. We believe that our findings suggest that Covid-19 has the potential to reshape global production networks and alter the way policymakers think about international dependencies in critical sectors. As

Covid Economics 53, 23 October 2020: 72-109

such our findings as well as the answers to these future research questions are relevant to private and public decision makers alike.

References

JaeBin Ahn, Mary Amiti, and David E Weinstein. Trade finance and the great trade collapse. *American Economic Review*, 101(3):298–302, 2011.

Carlo Altomonte, Filippo Di Mauro, Gianmarco Ottaviano, Armando Rungi, and Vincent Vicard. Global value chains during the great trade collapse: a bullwhip effect? *Firms in the international economy: Firm heterogeneity meets international business*, pages 277–308, 2012.

James E Anderson and Eric Van Wincoop. Gravity with gravitas: A solution to the border puzzle. *American economic review*, 93(1):170–192, 2003.

Mitsuyo Ando and Fukunari Kimura. How did the japanese exports respond to two crises in the international production networks? the global financial crisis and the great east japan earthquake. *Asian Economic Journal*, 26(3):261–287, 2012.

Natalie Bachas, Peter Ganong, Pascal J Noel, Joseph S Vavra, Arlene Wong, Diana Farrell, and Fiona E Greig. Initial impacts of the pandemic on consumer behavior: Evidence from linked income, spending, and savings data. Technical report, National Bureau of Economic Research, 2020.

Scott R Baker, Nicholas Bloom, Steven J Davis, and Stephen J Terry. Covid-induced economic uncertainty. Working Paper 26983, National Bureau of Economic Research, April 2020. URL http://www.nber.org/papers/w26983.

Richard Baldwin. The greater trade collapse of 2020: Learnings from the 2008-09 great trade collapse. *VoxEU.org. URL https://voxeu. org/article/greater-trade-collapse-2020*, 2020a.

Richard Baldwin. To treat covid-19's economic impact, start by keeping the lights on. *Chicago Booth Review*, 2020b.

Covid Economics 53, 23 October 2020: 72-109

Richard Baldwin and Rebecca Freeman. Supply chain contagion waves: Thinking ahead on manufacturing 'contagion and reinfection' from the covid concussion. *VoxEU.org. URL https://voxeu.org/article/covid-concussion-and-supply-chain-contagion-waves*, 2020.

Richard Baldwin and Eiichi Tomiura. Thinking ahead about the trade impact of covid-19. *Economics in the Time of COVID-19*, 59, 2020.

Richard E Baldwin. *The great trade collapse: Causes, consequences and prospects*. Cepr, 2009.

Richard Barichello. The covid-19 pandemic: Anticipating its effects on canada's agricultural trade. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 2020.

Christoph E Boehm, Aaron Flaaen, and Nitya Pandalai-Nayar. Input linkages and the transmission of shocks: firm-level evidence from the 2011 tōhoku earthquake. *Review of Economics and Statistics*, 101(1): 60–75, 2019.

Barthélémy Bonadio, Zhen Huo, Andrei A Levchenko, and Nitya Pandalai-Nayar. Global supply chains in the pandemic. Working Paper 27224, National Bureau of Economic Research, May 2020. URL http://www.nber.org/papers/w27224.

Tomaz Cajner, Leland D Crane, Ryan A Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz. The us labor market during the beginning of the pandemic recession. Technical report, National Bureau of Economic Research, 2020.

Bruno Casella, Richard Bolwijn, Daniel D Moran, and Keiichiro Kanemoto. Improving the analysis of global value chains: the unctad-eora database. *Transnational Corporations Journal*, 26(3), 2019.

Davin Chor and Kalina Manova. Off the cliff and back? credit conditions and international trade during the global financial crisis. *Journal of international economics*, 87(1):117–133, 2012.

Olivier Coibion, Yuriy Gorodnichenko, and Michael Weber. Labor markets during the covid-19 crisis: A preliminary view. Technical report, National Bureau of Economic Research, 2020.

Meredith Crowley and Xi Luo. Understanding the great trade collapse of 2008-09 and the subsequent trade recovery. *Economic Perspectives*, 35(2):44, 2011.

Covid Economics 53, 23 October 2020: 72-109

Ana Fernandes and Heiwai Tang. How did the 2003 sars epidemic shape chinese trade? *Covid Economics, Vetted and Real-Time Papers*, 22:154–176, 2020.

Keith Head and Thierry Mayer. Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics*, volume 4, pages 131–195. Elsevier, 2014.

International Monetary Fund. Policy Responses to Covid-19, 2020. URL https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19. Last accessed 04 August 2020.

Robert C Johnson and Guillermo Noguera. Accounting for intermediates: Production sharing and trade in value added. *Journal of international Economics*, 86(2):224–236, 2012.

Robert Koopman, Zhi Wang, and Shang-Jin Wei. Tracing value-added and double counting in gross exports. *American Economic Review*, 104(2):459–94, 2014.

Fernando Leibovici, Ana Maria Santacreu, et al. International trade of essential goods during a pandemic. *Covid Economics, Vetted and Real-Time Papers*, 21:59–99, 2020.

Andrei A Levchenko, Logan T Lewis, and Linda L Tesar. The collapse of international trade during the 2008–09 crisis: in search of the smoking gun. *IMF Economic review*, 58(2):214–253, 2010.

Man Li, Tao Ye, Peijun Shi, and Jian Fang. Impacts of the global economic crisis and tohoku earthquake on sino–japan trade: a comparative perspective. *Natural Hazards*, 75(1):541–556, 2015.

Josephine Ma. Coronavirus: China's first confirmed covid-19 case traced back to november 17. *South China Morning Post*, 2020. URL https://www.scmp.com/news/china/society/article/3074991/coronavirus-chinas-first-confirmed-covid-19-case-traced-back.

Maryla Maliszewska, Aaditya Mattoo, and Dominique Van Der Mensbrugghe. The potential impact of covid-19 on gdp and trade: A preliminary assessment, 2020.

Dimitris Papanikolaou and Lawrence D.W. Schmidt. Working remotely and the supply-side impact of covid-19. Working Paper 27330, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27330.

Covid Economics 53, 23 October 2020: 72-109

Yasuyuki Todo, Kentaro Nakajima, and Petr Matous. How do supply chain networks affect the resilience of firms to natural disasters? evidence from the great east japan earthquake. *Journal of Regional Science*, 55(2):209–229, 2015.

World Trade Organization. World Trade Statistical Review 2020, 2020a. URL https://www.wto.org/english/res_e/statis_e/wts2020_e/wts20_toc_e.htm. Last accessed 31 August 2020.

World Trade Organization. Trade falls steeply in first half of 2020, 2020b. URL https://www.wto.org/english/news_e/pres20_e/pr858_e.htm. Last accessed 10 October 2020.

Covid Economics 53, 23 October 2020: 72-109