Centre for Economic Performance

Discussion Paper ISSN 2042-2695

No. 1981
February 2024

Endogenous mobility in pandemics: Theory and evidence from the United States

Xiao Chen
Hanwei Huang
Jiandong Ju
Ruoyan Sun
Jialiang Zhang

LSE The London School of Economics and Political Science

UKRI Economic and Social Research Council

Abstract
We study infectious diseases in a spatial epidemiology model with forward-looking individuals who weigh disease environments against economic opportunities when moving across regions. This endogenous mobility allows regions to share risk and health resources, resulting in positive epidemiological externalities for regions with high R0s. We develop the Normalized Hat Algebra to analyze disease and mobility dynamics. Applying our model to US data, we find that cross-state mobility controls that hinder risk and resource sharing increase COVID-19 deaths and decrease social welfare. Conversely, by enabling "self-containment" and "self-healing," endogenous mobility reduces COVID-19 infections by 27.6% and deaths by 22.1%.

Keywords: SIRD model, spatial economy, endogenous mobility, basic reproduction number, Normalized Hat Algebra, containment policies
JEL Codes: C61; D91; F60; I12; I18; J61; R13

This paper was produced as part of the Centre's Trade Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

For helpful comments, we would like to thank Lorenzo Caliendo, Jingting Fan, Hanming Fang, Wenlan Luo, Yang Jiao, Steffan Qi, Larry Qiu, Chang Sun, and the seminar participants at numerous workshops, seminars, and conferences. We are responsible for any remaining errors. This project has received funding from the Research Grant Council of Hong Kong (project No. CityU 11501121).

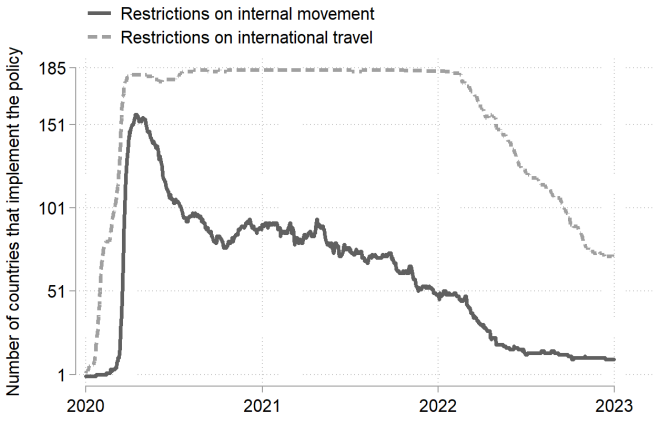
Xiao Chen, University of International Business and Economics. Hanwei Huang, City University of Hong Kong, CIFER and Centre for Economic Performance at London School of Economics. Jiandong Ju, Tsinghua University. Ruoyan Sun, University of Alabama. Jialiang Zhang, Central University of Finance and Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

The 2019 novel coronavirus (COVID-19) pandemic has caused an enormous loss of life worldwide. In response, governments around the world implemented widespread restrictions on internal movement and international travel (see Figure 1). These mobility control policies come with significant economic costs, leaving policymakers struggling to balance the health benefits and the economic burdens created by these controls. However, evaluating the costs and benefits of mobility control policies presents significant challenges. Their true policy effects are often intertwined with people's behavioral responses to infectious diseases (Fang et al., 2020; Goolsbee and Syverson, 2021). In addition, epidemiological processes exhibit highly non-linear dynamics, rendering estimations based on simplistic (log-) linear trajectories inadequate (Baldwin, 2020).

Figure 1: Number of Countries with COVID-19 Travel Restrictions



Data source: Oxford COVID-19 Government Response Tracker (OxCGRT).

In this paper, we build a spatial epidemiological model with forward-looking agents to address these challenges. The model captures people's behavioral responses to epidemiological conditions and accounts for the intrinsic non-linearity of infectious diseases. The model generates new insights into epidemiological externalities and enables policy evaluations through counterfactual simulations. Our key findings are as follows. First, although cross-region mobility exposes people to infectious diseases in other regions, which constitutes a negative epidemiological externality, endogenous mobility creates positive epidemiological externalities by allowing regions to share risk and medical resources. This works through both a "self-containment" effect, wherein susceptible people move away from regions with high infection rates to regions with low infection rates, and a "self-healing" effect, which occurs when infected people move to regions with higher recovery or lower death rates than other regions. Second, when applying our model to data from the United States (US), we find that, without cross-state mobility control policies, COVID-19 infections in

the US would have increased by 10.80%, but COVID-19 deaths would have dropped by 55.1%. By enabling the "self-containment" and "self-healing" effects, our model indicates that endogenous cross-region mobility reduces infections by 27.6% and deaths by 22.1% when compared to a model with exogenous mobility. Finally, our simulation results indicate that there may be no middle ground concerning an optimal mobility control policy. Policymakers may prefer to implement mobility control as strictly as possible in the face of fatal infectious diseases and as loosely as possible when dealing with highly transmissible but less deadly diseases like seasonal flu and COVID-19 Omicron. Overall, our analyses reveal the complex nature of cross-region mobility control policies and call for caution in their implementation.

The framework we develop is a multi-region Susceptible-Infected-Recovered-Deceased (SIRD) model that incorporates interactions between human mobility and epidemiological dynamics. It encompasses two fundamental mechanisms from the epidemiology and economic geography literature. *First*, the virus is transmitted by contact between susceptible and infected people. In the presence of human mobility, both local and cross-region contacts generate new infections.[1] We assume that the probability of infection (infection rate) depends on the density of susceptible and infected individuals, which varies endogenously across regions and over time. Infected individuals either recover or die, and the probability of recovery (recovery rate) and the probability of death (death rate) differ across regions.[2] *Second*, individuals are forward-looking when moving across regions. They account for mobility costs, which depend on containment policies, and for factors that determine expected utilities at their destinations (Artuç et al., 2010; Caliendo et al., 2019; Redding and Rossi-Hansberg, 2017). In the presence of infectious diseases, individuals trade off variations in infection, recovery, and death rates and economic opportunities across regions when choosing an optimal destination. For example, the expected utility of susceptible individuals is a weighted average of utilities of infected and susceptible people, with the weight determined by infection rates, discounted by the inter-temporal discount rate, and mobility costs. We find evidence supporting both model mechanisms using US data covering the COVID-19 pandemic.

The model makes novel predictions about mobility, disease dynamics, and containment policy. *First*, the model predicts that cross-region mobility depends on the disease environment and that it varies across individuals and over time. For example, we show that susceptible individuals' mobility rate declines with a higher infection rate at their destination. In other words, susceptible individuals avoid moving to high-risk areas. This prediction is consistent with cross-state mobility

[1] An extensive body of epidemiological literature incorporates spatial mobility into the SIRD model, which typically adopts a statistical approach to cross-region mobility (e.g., Balcan et al., 2009; Chinazzi et al., 2020). Some agent-based studies examine individual mobility decisions (e.g., Eubank et al., 2004; Longini Jr et al., 2005) but do not consider the forward-looking nature of mobility decisions.

[2] These variations may result from differences in medical resources across regions. Our model could be easily extended to incorporate time-variant recovery and death rates, but we do not have sufficient data to estimate them. We therefore assume these rates to be time-invariant.

patterns observed in the US during the COVID-19 pandemic.

Second, we show that endogenous cross-region mobility generates a positive epidemiological externality: regions share risk and medical resources, enabling "self-containment" and "self-healing" effects. We characterize properties of the basic reproduction number, i.e., the expected number of new cases generated by an infected individual at the outbreak (Allen and Van den Driessche, 2008; Diekmann et al., 2010), which is called the *global* R_0 in our setup with multiple regions (Antràs et al., 2023). We show that global R_0 falls between the maximum and minimum values of *local* R_0s, with the local R_0 representing the basic reproduction number of a region under mobility autarky. In contrast, the model by Antràs et al. (2023) predicts that the open-economy global R_0 exceeds the maximum local R_0s. They therefore posit that globalization may make pandemics more likely to occur because cross-region interactions expose people to foreign epidemics. In our model, regions further share risk and medical resources by allowing susceptible people to move away from regions with high infection rates (a "self-containment" effect) and allowing infected people to move to regions with high recovery or low death rates (a "self-healing" effect), both of which tend to reduce global infections and deaths.[3] Therefore, although the epidemiological externality negatively affects regions with low local R_0s (i.e., regions with low infection rates and high recovery rates), it positively affects regions with high local R_0s (i.e., regions with high infection rates and low recovery rates).

Third, we show that in the presence of cross-region mobility, global R_0 declines whenever transmission rates (probability of infection upon contact between infected and susceptible individuals) decrease or removal rates (the sum of death and recovery rates) increase in any region. The above result provides a rationale for prioritizing disease control in the region with the highest local R_0: if the pandemic can be controlled there, it can be controlled globally. This result suggests all regions matter for disease control when people can move across regions. Due to epidemiological externality, improved infectious disease control in any region contributes to global disease control.

Beyond examining R_0 and disease dynamics at the outbreak, we develop a new method – the "Normalized Hat Algebra" (NHA) – to study the full dynamics of the model. Existing methods to solve quantitative spatial models, such as the "Exact Hat Algebra" for static models (Dekle et al., 2007; Redding and Rossi-Hansberg, 2017) and the "Dynamic Hat Algebra (DHA)" for dynamic models (Caliendo et al., 2019), cannot be used for our model. The main difficulty is that infection rates depend on the distributions of individuals across regions, which change endogenously over time. This makes it impossible to follow Caliendo et al. (2019) to solve the model in terms of relative time differences.[4] To overcome this hurdle, we develop the NHA which normalizes vari-

[3] We view our analysis as complementary to Antràs et al. (2023). Their analysis fits the global context, as people typically cannot enjoy medical resources in other countries. Our analysis fits the domestic context, as people typically have access to medical resources when moving to other regions within the same country.

[4] Our model can be solved using DHA in the special case where infection rates are constant over time.

ables by their corresponding steady-state values and solves the model in terms of the normalized variables. Although this method requires knowledge of the steady-state equilibrium, it enables us to compute transitional dynamics and conduct counterfactual simulations without knowing all model fundamentals along the transitional path. Simulation results indicate that the epidemic and mobility dynamics are highly non-linear: there can be multiple waves of infections, and mobility flows between regions can overshoot or undershoot their steady-state levels.

Applying the NHA, we quantitatively evaluate COVID-19 containment policies in the US through the lens of an estimated model of 50 US states and Washington, DC. Relying on structural equations derived from the model, we first use COVID-19 epidemiological data and high-frequency mobility data from *PlaceIQ* (Couture et al., 2022) to estimate the effect of mobility control policies on cross-state mobility costs, and the effect of local containment policies on the transmission rate. Our estimations imply that a one-point increase in mobility control by a given state raises the cross-state mobility cost by about 1.60%, and a one-point increase in local containment policy decreases the transmission rate by about 0.30%.[5]

Armed with other model parameters estimated for 2020, including the steady-state equilibrium calibrated to match the pre-pandemic US economy, we conduct policy evaluations by simulating the estimated baseline economy and three counterfactual scenarios: *i.)* exogenous mobility, with cross-state mobility rates fixed at the pre-pandemic level; *ii.)* no cross-state COVID-19 mobility control policies; and *iii.)* no local COVID-19 containment policies. The main findings from our analysis are as follows. *First*, the comparison between the baseline economy and scenario *i* reveals that without endogenous mobility, there would have been 31.3 million more infections and 1.28 million more deaths. Intuitively, with endogenous mobility, there is a "self-containment" effect that deters people from moving to states with high infection rates. Additionally, there is a "self-healing" effect that encourages people to move to states with high recovery and low death rates. As both effects tend to reduce infections and deaths, it is vital not to impede them during a pandemic. *Second*, when comparing the baseline economy and scenario *ii*, we find that without cross-state mobility controls, there would have been 8.9 million more infections but 2.50 million fewer deaths. Mobility control policies reduce contact among people across states, thus leading to fewer infections, but they also dampen the self-healing effect and increase deaths. Due to higher death tolls and mobility costs, welfare declines in the presence of cross-state mobility control. *Third*, when examining scenario *iii*, we find that local containment policies effectively *flatten the epidemic curve*. Without such policies, infections would have risen more sharply, and an additional 62.8 million infections would have occurred. In addition, there would have been 1.33 million more

[5]Mobility control policies include travel recommendations and travel restrictions. Local containment policies include measures such as closures of schools, workplaces, and public transport. We follow Hale et al. (2021) in digitizing the policies imposed by US states. Appendix Table B.1 provides a detailed list of these policies.

deaths, close to the number from scenario *i*. We therefore find that endogenous mobility contributes greatly to saving lives, playing a role comparable to local containment policies.

We make several contributions to the literature. First, we contribute to the literature that bridges economics and epidemiological models. Building on compartmental models in epidemiology, existing studies examine quarantines and testing (Berger et al., 2020; Piguillem and Shi, 2022), information disclosure (Argente et al., 2022), lockdowns (Acemoglu et al., 2021; Alvarez et al., 2021; Fajgelbaum et al., 2021), and the economic impact of COVID-19 (Arellano et al., 2023; Atkeson, 2020; Eichenbaum et al., 2021). The studies most similar to ours are Argente et al. (2022), Bisin and Moro (2022), and Antràs et al. (2023), which also explore spatial interactions during a pandemic. The unique feature of our framework is that we model heterogeneous and forward-looking agents who optimize their mobility decisions. While Argente et al. (2022) assume that susceptible, recovered, and infected people all have the same mobility rate, we consider the different mobility problems they face. In addition, in their model, mobility reacts to the pandemic only via mobility costs, which they assume to depend directly on the number of cases disclosed. In our model, the pandemic affects mobility by shaping people's expected utility. Bisin and Moro (2022) study within-city mobility and assume that individuals follow exogenous behavioral rules, whereas our modeled individuals optimize their cross-region mobility choices. Finally, Antràs et al. (2023) consider spatial interactions via international trade without people changing locations while we study human mobility explicitly.[6] As mentioned above, this leads to very different predictions on the global R_0 and epidemiological externality.

We also contribute to the quantitative economic geography literature (Allen and Arkolakis, 2014; Caliendo et al., 2019; Kleinman et al., 2023; Monte et al., 2018; Redding and Rossi-Hansberg, 2017; Tombe and Zhu, 2019). The key innovation of our model is that individuals can switch types and face different mobility problems in subsequent periods. The switching probability is time-variant, endogenous, and depends on the distribution of individuals. This makes existing quantitative methods for dynamic mobility models, e.g., Caliendo et al. (2019), unsuitable for our study.[7] Thus, we also make a methodological contribution by developing the NHA, which can be used to solve other models similar to ours. For example, our method can be applied to study spatial labor markets wherein workers search for jobs across markets and can switch between being employed and unemployed, the probability of which changes endogenously over time.[8]

[6] Similar to Antràs et al. (2023), Giannone et al. (2022) also consider the spatial diffusion of infectious disease via trade. They adopt a reduced-form relationship between trade and mobility flows.

[7] Caliendo et al. (2021) extend the study of Caliendo et al. (2019) to models with multiple types of agents. They study the mobility of workers with different, but fixed, skills.

[8] There are studies that examine job searches across space, although in a static environment (e.g., Manning and Petrongolo, 2017). There is also an established body of literature on dynamic job searches (e.g., Rogerson et al., 2005), which typically ignores mobility across space. Recent research has begun modeling spatial mobility dynamics but with a focus on the steady state (Schmutz and Sidibé, 2019).

Finally, we contribute to the literature on infectious disease policies. There is substantial evidence that non-pharmaceutical policies are effective in containing viruses and flattening epidemic curves (Hsiang et al., 2020; Maier and Brockmann, 2020) and that these policies have nuanced welfare implications (Adda, 2016; Argente et al., 2022; Farboodi et al., 2021; Giannone et al., 2022). Our framework allows us to separate policy effects from behavioral responses, accounting for the intrinsically non-linear process of disease dynamics. Consistent with previous findings (Fang et al., 2020; Farboodi et al., 2021; Goolsbee and Syverson, 2021), we find a substantial behavioral effect in reducing infections and deaths. Different from these studies, we emphasize that there is a "self-healing" effect that prompts infected people to move to high-recovery and low-death regions, other than a "self-containment" effect, which induces susceptible people to move away from regions with high infection risk. While the latter mainly reduces infections, the former lowers the death toll. In addition, when we examine optimal cross-region mobility control policies, we find that the long-run optimal cross-region mobility control policy may have no middle ground: a benevolent policymaker who optimizes aggregate welfare across regions prefers mobility controls that are either as strict as possible or as loose as possible. This finding differs from previous studies, which tend to find interior solutions for optimal policies (Acemoglu et al., 2021; Fajgelbaum et al., 2021; Giannone et al., 2022). If an infectious disease becomes less deadly or more transmittable, a policymaker is more likely to prefer loose mobility controls, which may rationalize the loosening of COVID-19 mobility control policies over time, as observed in Figure 1.

The rest of the paper is organized as follows. Section 2 presents the stylized facts that motivate our model. Section 3 sets up the model and discusses analytical results on R_0 and mobility. Section 4 introduces the NHA and presents simulation findings on disease and mobility dynamics. Section 5 applies the model to the US, calibrates it with the data, and conducts policy evaluations by counterfactual simulations. Section 6 discusses the optimal mobility control policy and the extensions and limitations of our model. Section 7 concludes the paper.

2 Motivating Stylized Facts

This section provides three stylized facts about pandemics, containment policies, and human mobility. We first show how COVID-19 containment policies and infection risks vary across regions and change over time. We then present two facts on the interaction between cross-region human mobility and the COVID-19 pandemic. These facts motivate our model in the subsequent section.

2.1 Data

These facts are generated using three COVID-19 datasets, which will also be used to quantify our theoretical model later.

Epidemiological Data Our COVID-19 epidemiological data are from the *COVID Tracking Project* (Miller and Curry, 2020). We collect the daily number of confirmed cases, recovery cases, and deaths for each US state. We aggregate the daily data to build a panel dataset at the state and bi-week levels from January 2020 to August 2021.

Containment Policy Data The *OxCGRT*, i.e., Oxford Covid-19 Government Response Tracker (Hale et al., 2021), collects COVID-19 containment policies across the world at national and sub-national level. From OxCGRT, we obtained 8 policy measures at the state level under the containment and closure policies category for the US, which are listed as C1-C8 in appendix Table B.1. We classify these 8 measures into two types: local containment policies and cross-region mobility control policies. Local containment policies include closings of schools/universities, workplaces, and public transport, cancellations of public events, and orders to stay home (C1-C6). Cross-region mobility control policies include restrictions on internal movement between cities/regions (C7) and restrictions on international travel (C8). Numbers are assigned to each measure to capture its stringency. For example, the mobility restriction indicator equals 0 when the local government implements no restrictions, 1 when the government recommends against traveling across regions, and 2 when cross-region movement is restricted. Following Hale et al. (2021), we construct state-level policy indexes by first re-scaling each of the policy measures by its maximum value to create a score between 0 and 100, and then creating composite policy indicators for each state by averaging across the sub-policy indexes.

Mobility Data We use anonymous smartphone location data from *PlaceIQ* to measure human mobility. *PlaceIQ* data tracks mobile phone users' location over time and records smartphone devices "pinging" in a given geographic unit on a given day. Using this data, Couture et al. (2022) compute a daily measure of location exposure index (LEX), which is the fraction of active phones in a geographic unit on a day that was active in another geographic unit at any point in the last 14 days. We get the daily state-to-state LEX measure and aggregate it to a biweekly frequency in two ways: 1). the average within each bi-week, and 2). the maximum within each bi-week.

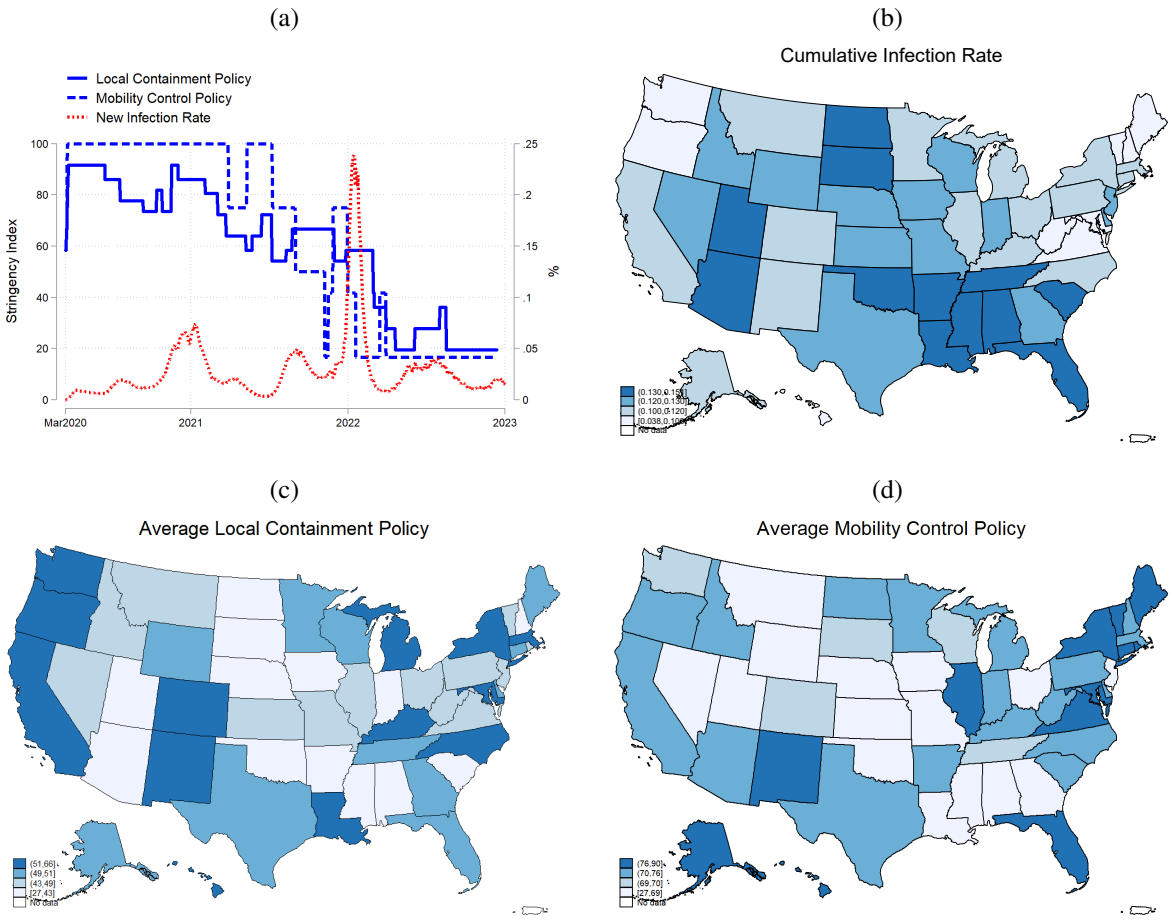
2.2 Results

Fact 1. *Infection rates and COVID-19 containment policies vary significantly across regions and over time.*

Figure 2 (a) shows the evolution of the US COVID-19 infection rate, measured by the number of new cases within each bi-week divided by the US population.[9] As we can see, the infection rate varies over time in a highly non-linear way: there were multiple waves of infection. Figure 2 (b) plots the cumulative infection rate for each state, measured by the number of cumulative infected

[9]We obtain the national and state-level population data from the 2020 United States Census.

Figure 2: COVID-19 Pandemic and Containment Policies in the US



Notes: Figure (a) plots the stringency of local containment and mobility control policies averaged across US states and the national infection rate. Figure (b) plots cumulative infection rates: cumulative infected cases (till August 2021) over the state population. Figures (c) and (d) plot the average local containment and mobility control policy indexes from January 2020 to August 2021.

cases divided by the state population, which varies significantly across US states. The inland and southern states tend to have higher cumulative infection rates than the northern and coastal states.

COVID-19 containment policies also demonstrate substantial time and spatial variations. Figure 2 (a) also shows the local containment policies (C1-C6) and mobility control policies averaged across states (C7 and C8). We find that state governments imposed stringent local containment and cross-state mobility policies at the outbreak of the pandemic. Over time, they gradually relaxed these policies. Figure 2 (c) and (d) plot the average local containment and mobility control policies for each state, which differ substantially across states. Overall, inland states tend to adopt more lenient containment policies than coastal states.

Table 1: Cross-State Mobility and the Number of New Cases

	(1)	(2)	(3)	(4)
Dependent variable	ln (number of new cases)			
$\sum_{j\neq i} m_{ji}^{max} ln(NewCases_{j,t-1})$	0.374***			
	(22.695)			
$\sum_{j\neq i} m_{ji}^{mean} ln(NewCases_{j,t-1})$		0.394***		
		(23.017)		
$\sum_{j\neq i} m_{ji}^{max} ln(CumulativeCases_{j,t-1})$			0.159***	
			(13.737)	
$\sum_{j\neq i} m_{ji}^{mean} ln(CumulativeCases_{j,t-1})$				0.171***
				(13.973)
$Stringency\ Index_{t-1}$	-0.004*	-0.004*	-0.002	-0.002
	(-1.766)	(-1.742)	(-0.629)	(-0.623)
Observations	2,040	2,040	2,040	2,040
R-squared	0.971	0.972	0.943	0.943
BiWeek FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes

Notes: This table estimates the effect of interstate mobility on state-level COVID-19 cases in the US. The dependent variable is the logarithm of the number of new cases in state i during a two-week period. We use four measures of local exposure to outside cases caused by interstate mobility (columns 1-4). They combine two measures of initial interstate mobility and two measures of outside cases. The two measures of initial bilateral mobility are m_{ji}^{max} and m_{ji}^{mean}, the maximum and average daily mobility rate from state j to state i in the initial period, respectively. Two measures of outside cases are used: $ln(NewCase_{j,t-1})$ and $ln(CumulativeCase_{j,t-1})$, the logarithm of the lagged number of new cases and cumulative number of cases in origin state j, respectively. The $Stringency\ Index_{t-1}$ measures the stringency of closure policies, ranging from 0 to 100 at the state and biweekly levels. The numbers in parentheses are robust t-statistics with standard errors two-way clustered at the state and biweekly levels. Significance is indicated by *, **, and *** at the 0.1, 0.05, and 0.01 levels, respectively.

Fact 2. *People fleeing from areas with infectious disease outbreaks tend to increase the number of cases in their destinations.*

Considering the impact of mobility on pandemics, an intuitive question is whether having people moving from states with more cases leads to an increase in local cases.[10] To examine that, we first merge the three datasets discussed above to form a panel covering 50 US states and DC at the cross-section across 40 bi-weeks from January 2020 to August 2021. The summary statistics of the data are shown in appendix Table B.2. Panel A is at the state-pair biweekly level, while Panel B is at the biweekly state level. Using the data, we estimate the following equation:

$$ln(y_{i,t}) = a_0 + a_1 \sum_{j\neq i} m_{ji}^0 \ln(cases_{j,t-1}) + a_2 StringencyIndex_{i,t-1} + d_t + S_i + \nu_{i,t}, \quad (1)$$

where $y_{i,t}$ is the number of newly confirmed COVID-19 cases in state i at a bi-week t, $Stringency\ Index_{i,t-1}$ captures the lagged average containment and cross-region mobility control policies of

[10]Similarly, Glaeser et al. (2022) study the effect of within-city mobility changes on local cases at zip code levels.

state i,[11] d_t is the time-fixed effect that absorbs time-varying shocks at the national level, and S_i is the state-fixed effect that absorbs state-specific time-invariant factors. $\sum_{j \neq i} m_{ji}^0 \ln(cases_{j,t-1})$ measures the exposure of state i to the pandemic in other states due to inter-state human mobility. m_{ji}^0 is the share of people moving from state j to state i in the initial period of our sample, which is the second bi-week of January 2020, before the outbreak of the pandemic. We use two measures to measure m_{ji}^0: the maximum daily mobility flows during a bi-week period, and the average daily mobility flows. We use $cases_{j,t-1}$ to capture the severity of the pandemic at an origin state j, of which there are two measures: the number of new cases in period $t-1$, and the number of cumulative cases by period $t-1$. Interacting the two measures of $cases_{j,t-1}$ with the two measures of m_{ji}^0, we have four approaches to measuring a state i's exposure to the pandemic in other states, which corresponds to the four columns in Table 1. Across all columns, we find that having incomers from states with more cases is significantly associated with more local cases. The effect of mobility exposure weighted in terms of the lagged number of new cases appears to be larger than the cumulative number of cases.

Fact 3. *People avoid moving to areas with high infection rates.*

The above result indicates that mobility has a strong effect on the pandemic. A natural follow-up question is whether mobility responds to the pandemic. To check that, we examine interstate mobility flows in the US before and during the COVID-19 pandemic using the following empirical specification:

$$\begin{aligned} m_{ij,t} &= b_0 + b_1 Infection_{i,t-1} + b_2 Infection_{j,t-1} + b_3 MobilityRestriction_{i,t-1} \\ &+ b_4 MobilityRestriction_{j,t-1} + \sum\nolimits_k \beta_k X_{ij,k} + d_t + O_i + D_j + \xi_{ij,t}, \end{aligned} \tag{2}$$

where $m_{ij,t}$ is the number of state i people moving to state j in a bi-week t relative to state j's population. The key variables of interest are $Infection_{i,t-1}$ and $Infection_{j,t-1}$, which captures infection rates at the origin state and destination state in the lagged period. We control for $MobilityRestriction_{i,t-1}$ and $MobilityRestriction_{j,t-1}$, which are mobility restrictions (see C7 in Table B.1) imposed by the origin and destination states, respectively. We also include gravity variables ($X_{ij,k}$), time fixed effect (d_t), and origin and destination fixed effects (O_i and D_j). We estimate the model using Poisson Pseudo Maximum Likelihood (PPML), to deal with zeros in the mobility data and potential heteroskedasticity (Silva and Tenreyro, 2006).

The results are presented in Table 2. Columns (1) and (2) use the maximum daily mobility flows during each bi-week as the dependent variable, and columns (3) and (4) use the average during the same period. Across all columns, we find that higher infection rates at the destination

[11] It is the average of C1 - C8 in Table B.1. Lagged variables are used to ease potential endogeneity concerns that containment policies could be affected by the severity of the disease in a region.

Table 2: Infection Risk and State-to-State Mobility

	(1)	(2)	(3)	(4)
Dependent variable	Max mobility		Average mobility	
new infection rate of origin	-0.940	-0.491	-0.601	-0.146
	(-0.436)	(-0.642)	(-0.272)	(-0.201)
new infection rate of destination	-7.633***	-7.844***	-7.456***	-7.698***
	(-3.432)	(-10.862)	(-3.245)	(-11.085)
mobility restrictions of origin	-0.048***	-0.048***	-0.048***	-0.048***
	(-3.531)	(-9.873)	(-3.408)	(-9.602)
mobility restrictions of destination	-0.007	-0.008	-0.003	-0.005
	(-0.448)	(-1.498)	(-0.201)	(-0.827)
log(distance)	-1.364***		-1.384***	
	(-271.5)		(–214.0)	
shared state border	1.069***		1.100***	
	(95.5)		(96.9)	
Observations	102,000	102,000	102,000	102,000
Time FE	Yes	Yes	Yes	Yes
Origin FE	Yes	No	Yes	No
Destination FE	Yes	No	Yes	No
Origin-Destination Pair FE	No	Yes	No	Yes

Notes: The dependent variables are the maximum daily bilateral mobility flow within a bi-week in columns (1)-(2) and the average in columns (3)-(4). Independent variables $new\ infection\ rate\ of\ origin$ and $new\ infection\ rate\ of\ destination$ are the number of new cases divided by population in the origin and destination states, respectively. The models are estimated using Poisson Pseudo Maximum Likelihood (PPML). The numbers in parentheses are robust t-statistics with standard errors two-way clustered at the state pair and biweekly levels. Significance is indicated by *, **, and *** at the 0.1, 0.05, and 0.01 levels, respectively.

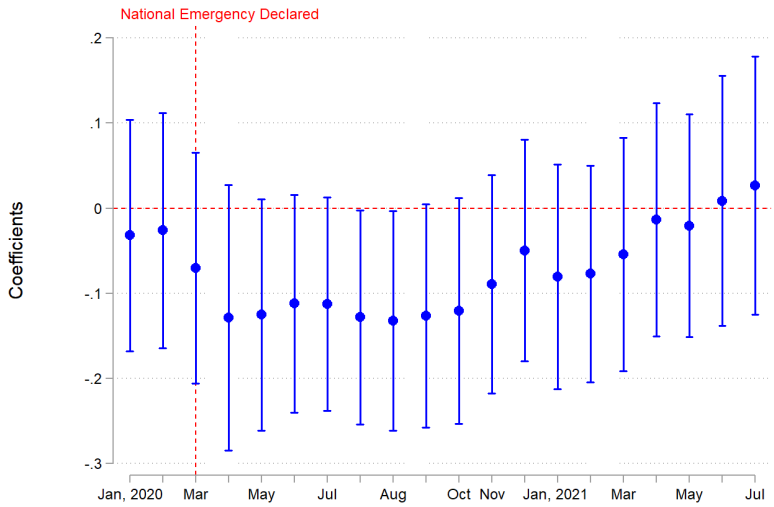
state are associated with lower mobility inflows. A higher infection rate at the origin state is also associated with lower mobility flows, albeit insignificant. Mobility restrictions also appear to have negative effects on mobility flows.

There may be concerns that lagged infection rates do not fully reflect infection risk, bilateral mobility flows do not capture overall mobility inflows of a region, and the above estimations only capture the average effect. To deal with these concerns, we also examine the effect of cumulative infection rates on total net mobility inflows using the following Difference-in-Differences specification with time-varying treatment effects:

$$\begin{aligned} NetInflowShare_{j,t} &= a_0 + \sum\nolimits_k a_k \cdot CumulativeInfectionRate_j \times \tau_{k,t} \\ &+ c_2 MobilityRestriction_{j,t-1} + d_t + S_j + \nu_{j,t}, \end{aligned} \tag{3}$$

where $NetInflowShare_{j,t}$ is the net total mobility inflows from other states to a state j at period t divided by the population of j, $CumulativeInfectionRate_j$ is the cumulative infection rate of state j measured by the total number of confirmed cases at the end of the sample divided its

Figure 3: Effect of Cumulative Infection Rates on Net Mobility Inflows



Notes: This figure depicts the effect of cumulative infection rate on net mobility inflows of US states before and after the national emergency declaration in March 2020. The dependent variable is the net biweekly mobility inflows of a state, measured by the maximum daily net inflows within a bi-week. The markers represent the estimated coefficients of interaction terms, a_k, in equation (3). The reference group is the final month of the data sample, August 2021. The capped spikes represent the associated 95% confidence intervals, with standard errors clustered at the state and biweekly levels.

population, and $\tau_{k,t}$ is a time dummy that equals 1 if period t falls within a specific year-month k.

Figure 3 plots the estimated a_k, the coefficient of interaction terms of the time dummies and cumulative infection rates. We find that before the declaration of national emergency in March 2020, there were no significant differences in net mobility inflows across states. But as the national emergency was declared, net mobility inflows to states with high cumulative infection fell relative to states with low cumulative infection rates. Such differences in net inflows petered out as infection rates gradually declined in the first half of 2021 (see Figure 2 a).

3 Model

This section builds a spatial model of pandemic and endogenous human mobility. Given Fact 1, it is important to have a model featuring rich geographic heterogeneity in disease environments. Given Facts 2 and 3, the model should allow the pandemic to diffuse via human mobility, and human mobility to react to the pandemic.

The model we build is a multi-region SIRD model with forward-looking agents who make endogenous cross-region mobility decisions. Individuals consider the risk of infection and probabilities of recovery and death in different regions, which affect their expected utility when moving.

We will use the model to characterize disease and mobility dynamics in this section and do quantitative analyses in subsequent sections.

3.1 Mobility, Demographic, and Disease Dynamics

Let us consider an economy consisting of N regions. We follow the tradition of SIRD models (Kermack and McKendrick, 1927) to divide the total population into four mutually exclusive types: Susceptible (S), Infected (I), Recovered (R), and Deceased (D). We assume that time is discrete and use $\overline{S}_{i,t}, \overline{I}_{i,t}, \overline{R}_{i,t}, \overline{D}_{i,t}$ to denote the population of each type in region i at the beginning of period t, and $S_{i,t}, I_{i,t}, R_{i,t}, D_{i,t}$ the end of period t. The size of each type in each region changes due to both the pandemic and mobility across regions.

For mobility, we follow the recent quantitative economic geography literature (Caliendo et al., 2019; Redding and Rossi-Hansberg, 2017; Tombe and Zhu, 2019) to assume that individuals of each type have idiosyncratic preference shocks for locations drawn from extreme value distributions, an approach commonly adopted in discrete choice models (Anderson et al., 1992). After observing the preference shock, an individual moves to the region that yields the highest expected utility net of the mobility cost. Specifically, the problem faced by type S in region i in period t can be characterized by the following Bellman equation,

$$U^S_{i,t}(\boldsymbol{\varepsilon_{i,t}}) = u_i + \max_{j=1,...,N} \{\beta E_t[(1-\alpha_{j,t+1})U^S_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}}) + \alpha_{j,t+1}U^I_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}})] - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\}, \tag{4}$$

where u_i is the instantaneous utility of individuals in region i, and $\beta \in (0,1)$ is the inter-temporal discount rate. Conditional on moving to region j, with probability $\alpha_{j,t+1}$, the individual gets infected and joins type I whose value function is $U^I_{j,t+1}$. With probability $1-\alpha_{j,t+1}$, the individual remains uninfected and has a value function of $U^S_{j,t+1}$. Following Antràs et al. (2023) and Farboodi et al. (2021), we assume that agents form rational expectations about $\alpha_{j,t+1}$. Although an individual is uncertain about whether or not he/she will be infected in the next period, the aggregate risk of infection is determined by collective decisions and taken as given by each individual. The cost of moving from region i to region j in period t is $\widetilde{\mu}_{ij,t}$. Individuals currently in region i receive a vector of preference shocks $\boldsymbol{\varepsilon_{i,t}} \equiv \{\varepsilon_{i1,t}, ..., \varepsilon_{iN,t}\}$ at the end of each period with each element $\varepsilon_{ij,t}$ following an *i.i.d.* Gumbel distribution

$$\Pr\{\varepsilon_{ij,t} \leq \varepsilon\} = \exp\left\{-\exp\left\{-\varepsilon/\kappa - \gamma^{Euler}\right\}\right\}, \tag{5}$$

where κ is a parameter that controls the dispersion of a shock and γ^{Euler} is the Euler constant ensuring that the shock has zero mean. These shocks capture the reality that individuals have idiosyncratic reasons for traveling to each region.

Similarly, for types I and R, their value functions are given by

$$U^I_{i,t}(\boldsymbol{\varepsilon_{i,t}}) = u_i + \max_{j=1,\ldots,N}\{\beta E_t[(1-\gamma^R_j-\gamma^D_j)U^I_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}}) + \gamma^R_j U^R_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}}) + \gamma^D_j U^D] - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\}, \tag{6}$$

and

$$U^R_{i,t}(\boldsymbol{\varepsilon_{i,t}}) = u_i + \max_{j=1,\ldots,N}\{\beta E_t[U^R_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}})] - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\}, \tag{7}$$

respectively, where γ^R_j and γ^D_j are the recovery and death rates in region j, and U^D is the utility of the deceased, which we normalize as a constant. To keep the model non-trivial, we assume that $0 \leq \gamma^R_j < 1$, $0 \leq \gamma^D_j < 1$, and $\gamma^R_j + \gamma^D_j < 1$. If $\gamma^D_j = 0$, our model collapses to a SIR model.

Solving problems (4), (6), and (7), we have the following results.

Proposition 1. *The expected lifetime utility of types S, I, and R in region i at time t can be written recursively as*

$$\begin{aligned}
V^S_{i,t} &= \exp\left(\frac{u_i}{\kappa}\right)\sum\nolimits_{j=1}^{N}\left(V^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})}\left(V^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}}\left(\mu_{ij,t}\right)^{-1}, \\
V^I_{i,t} &= \exp\left(\frac{u_i}{\kappa}\right)\sum\nolimits_{j=1}^{N}\left(V^I_{j,t+1}\right)^{\beta(1-\gamma^R_j-\gamma^D_j)}\left(V^R_{j,t+1}\right)^{\beta\gamma^R_j}\left(V^D\right)^{\gamma^D_j}\left(\mu_{ij,t}\right)^{-1}, \\
V^R_{i,t} &= \exp\left(\frac{u_i}{\kappa}\right)\sum\nolimits_{j=1}^{N}\left(V^R_{j,t+1}\right)^{\beta}\left(\mu_{ij,t}\right)^{-1},
\end{aligned} \tag{8}$$

where $V^g_{i,t} = \exp(E_{t-1}(U^g_{i,t})/\kappa), g \in \{S, I, R\}$ and $V^D = exp(U^D/\kappa)$ is the value of type D, and $\mu_{ij,t} = \exp(\tilde{\mu}_{ij,t}/\kappa)$.

Proof. See appendix A.1.1. □

Equation (8) indicates that the expected lifetime utility of individuals depends on the contemporaneous utility and the option value of changing regions and types in the next period. The option value sums the expected values across regions, which are discounted by the mobility cost and the inter-temporal discount rate. The expected value of a region is a weighted geometric average of values across types, with the weights given by probabilities of being respective types.

We next obtain the mobility rate of each type of individual, summarized as follows.

Proposition 2. *The probabilities of types S, I, and R moving from region i to region j in period t are given by*

$$\begin{aligned}
m^S_{ij,t} &= \frac{\left(V^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})}\left(V^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}}\left(\mu_{ij,t}\right)^{-1}}{\sum_{k=1}^{N}\left(V^S_{k,t+1}\right)^{\beta(1-\alpha_{k,t+1})}\left(V^I_{k,t+1}\right)^{\beta\alpha_{k,t+1}}\left(\mu_{ik,t}\right)^{-1}}, \\
m^I_{ij,t} &= \frac{\left(V^I_{j,t+1}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)}\left(V^R_{j,t+1}\right)^{\beta\gamma^R_j}\left(V^D\right)^{\beta\gamma^D_j}\left(\mu_{ij,t}\right)^{-1}}{\sum_{k=1}^{N}\left(V^I_{k,t+1}\right)^{\beta(1-\gamma^R_k-\gamma^D_k)}\left(V^R_{k,t+1}\right)^{\beta\gamma^R_k}\left(V^D\right)^{\beta\gamma^D_k}\left(\mu_{ik,t}\right)^{-1}},
\end{aligned} \tag{9}$$

$$m_{ij,t}^{R} = \frac{\left(V_{j,t+1}^{R}\right)^{\beta}\left(\mu_{ij,t}\right)^{-1}}{\sum_{k=1}^{N}\left(V_{k,t+1}^{R}\right)^{\beta}\left(\mu_{ik,t}\right)^{-1}}.$$

Proof. See appendix A.1.2. □

There are a few important differences between the above expressions and existing models. First, the mobility rates vary across types, whereas existing spatial models typically ignore mobility heterogeneity by assuming workers or commuters share the same mobility rate (e.g., Argente et al., 2022).[12] Second, individuals who can switch types (types S and I) consider that they might switch to other types in the following period, whereas individuals of the terminated type R do not. Finally, in addition to mobility friction, disease conditions (infection, recovery, and death rates) directly affect the mobility rate. In particular, the susceptible individuals' mobility rates depend on infection rates $\alpha_{j,t+1}$, which varies endogenously over time.

As is standard in SIRD models (Allen and den Driessche, 2008), demographic dynamics within a period caused by the disease is given by

$$\begin{aligned} S_{i,t} &= \overline{S}_{i,t} - T_{i,t} \\ I_{i,t} &= T_{i,t} + \left(1 - \gamma_i^R - \gamma_i^D\right)\overline{I}_{i,t}, \\ R_{i,t} &= \overline{R}_{i,t} + \gamma_i^R \overline{I}_{i,t} \\ D_{i,t} &= \overline{D}_{i,t} + \gamma_i^D \overline{I}_{i,t}, \end{aligned} \tag{10}$$

where $T_{i,t}$ is the number of newly infected individuals in region i in period t. Following the literature (e.g., Acemoglu et al., 2021), we assume that it depends on the densities of susceptible and infected individuals in the local population, and satisfies

$$\frac{T_{i,t}}{\overline{L}_{i,t}} = \min\{\frac{\overline{S}_{i,t}}{\overline{L}_{i,t}}, \chi_{i,t}\frac{\overline{S}_{i,t}}{\overline{L}_{i,t}}\frac{\overline{I}_{i,t}}{\overline{L}_{i,t}}\}, \tag{11}$$

where $\overline{L}_{i,t} = \overline{S}_{i,t} + \overline{I}_{i,t} + \overline{R}_{i,t}$ is the local population and $\chi_{i,t}$ is the transmission rate, the probability that the virus transmits from type I to type S conditional on contact between them. The *min* operator is used to cover the case that all susceptible individuals are infected. Given the number of newly infected individuals, the infection rate of type S is

$$\alpha_{it} \equiv \frac{T_{i,t}}{\overline{S}_{i,t}}. \tag{12}$$

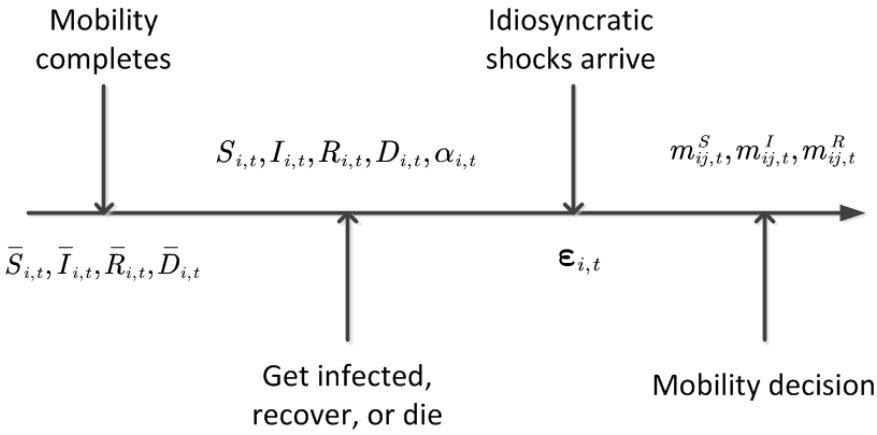
[12] A notable exception is Caliendo et al. (2021), who study the migration of workers with different skill types. However, they continue to assume that workers do not change types.

Last, the demographic dynamics across periods due to mobility are given by

$$\begin{aligned} \overline{S}_{i,t} &= \sum_{j=1}^{N} S_{j,t-1} m^{S}_{ji,t-1}, \ \overline{I}_{i,t} = \sum_{j=1}^{N} I_{j,t-1} m^{I}_{ji,t-1}, \\ \overline{R}_{i,t} &= \sum_{j=1}^{N} R_{j,t-1} m^{R}_{ji,t-1}, \ \overline{D}_{i,t} = D_{i,t-1}. \end{aligned} \tag{13}$$

The timing of events in the model is summarized in Figure 4.

Figure 4: Model Timeline



Notes: This figure plots the timeline of events in each model in period t. At the beginning of the period (after mobility in the previous period is complete), the stock of type g individuals in the region i is $\overline{g}_{i,t}, g \in \{S, I, R, D\}$. Next, possible infection (with probability $\alpha_{i,t}$), recovery, and death take place. The stock of type g individuals becomes $g_{i,t}$. At the end of the period, individuals derive their idiosyncratic preference shock $\varepsilon_{i,t}$ and make a mobility decision. The mobility rate of type g is $m^{g}_{ij,t}$.

3.2 Equilibrium

The model fundamentals are bilateral mobility costs $\mu_t \equiv \{\mu_{ij,t}\}_{i=1,j=1}^{N,N}$, the recovery rate $\gamma^R \equiv \{\gamma_i^R\}_{i=1}^{N}$, the death rate $\gamma^D \equiv \{\gamma_i^D\}_{i=1}^{N}$, the transmission rate $\chi_t \equiv \{\chi_{i,t}\}_{i=1}^{N}$, the value of type D, i.e., V^D, the dispersion parameter of the Gumbel distribution κ, and the discount rate β. We denote the time-varying fundamentals by $\Omega_t \equiv \{\mu_t, \ \chi_t\}$ and time-invariant fundamentals by $\Omega \equiv \{\gamma^R, \ \gamma^D, \ V^D, \ \kappa, \ \beta\}$.

We use $G_t \equiv \{S_{i,t}, I_{i,t}, R_{i,t}, D_{i,t}\}_{i=1}^{N}$ to denote distribution of individuals across regions, $m_t = \{m^S_{ij,t}, \ m^I_{ij,t}, \ m^R_{ij,t}\}_{i=1,j=1}^{N,N}$ their mobility rates, and $V_t = \{V^S_{i,t}, \ V^I_{i,t}, \ V^R_{i,t}\}_{i=1}^{N}$ their expected utility for each period, and define the sequential competitive equilibrium as follows.

Definition 1. *Given Ω, $\{\Omega_t\}_{t=0}^{\infty}$, and G_0, the sequential competitive equilibrium is a sequence of $\{G_t, \ m_t, \ V_t\}_{t=0}^{\infty}$ satisfying equations (8), (9), (10), (11), (12), and (13).*

We next define the *steady-state equilibrium*.

Definition 2. *A steady-state equilibrium is a sequential competitive equilibrium such that* $\{G_t,\ m_t,\ V_t\}_{t=0}^{\infty}$ *is constant over time.*

Finally, following the literature on the basic reproduction number R_0 (Van den Driessche, 2017), we define the *disease-free equilibrium* (DFE) as

Definition 3. *A DFE is defined as the point at which the whole population is susceptible, i.e.,* $S_{i,t}/L_{i,t} = 1,\ \forall i$.

In essence, the sequential competitive equilibrium characterizes the full dynamic path of the economy, while the steady-state equilibrium describes the economy once the pandemic ends, and the disease-free equilibrium represents the economy at the outbreak.

3.3 Mobility and Infection Risk

This subsection presents analytical results on mobility dynamics. SIRD models have analytical solutions only under certain restrictions (Harko et al., 2014). Incorporating them into an economic model with endogenous mobility makes it even more challenging to get analytical solutions. We first obtain results on the mobility of infected and recovered individuals.

Proposition 3. *If mobility costs are time-invariant, the value functions of types* I *and* R *are time-invariant, that is,* $V_{i,t}^R = V_i^R$ *and* $V_{i,t}^I = V_i^I$*; so are their mobility rates, i.e.,* $m_{ij,t}^R = m_{ij}^R$ *and* $m_{ij,t}^I = m_{ij}^I$, $\forall i\ \&\ j$.

Proof. See appendix A.1.3. □

Types I and R have both been infected. They are not concerned about the risk of infection and do not respond to changes in infection rates. Only type S individuals, who have not been infected, care about the risk of infection when making mobility choices. Therefore, without changes in mobility costs, mobility dynamics during a pandemic are shaped mostly by type S individuals.

Before discussing how type S responds to infection risk, we rank the value functions of agents. We assume that the value of the deceased is sufficiently small such that their value is always less than the value of recovered individuals, i.e., $V^D < V_{i,t}^R$.

Lemma 1. *If mobility costs are time-invariant, then we have: (a)* $V_i^R > V_i^I$ *and* $V_i^R \geq V_{i,t}^S$*; (b) near the DFE, we have* $V_i^I < V_{i,t}^S$, $\forall i$.

Proof. See appendix A.1.4. □

Result (a) implies that the value function of type R individuals is higher than those of types I and S. The intuition is that type I individuals might die while type S individuals might become infected and die, but type R has immunity and will not die. Near the DFE, as the infection rate approaches 0, type S's value approaches type R's value according to equation (8). Naturally, type S's value function tends to dominate type I's value function. Based on this result, we obtain the following proposition on how type S's mobility responds to the risk of infection.

Proposition 4. *Near the DFE, if the infection rate in region j increases, the mobility rate of type S to region j decreases:*

$$\frac{\partial \ln(m^S_{ij,t})}{\partial \alpha_{j,t+1}} = \beta(1 - m^S_{ij,t}) \ln(\frac{V^I_j}{V^S_{j,t+1}}) < 0. \tag{14}$$

Proof. See appendix A.1.5. □

This result implies that type S individuals avoid moving to regions with higher risks of infection, which is consistent with Fact 3 in Section 2. We also note that type S's mobility response is proportional to the (log) difference of their value function from type I's: $\ln(V^I_j) - \ln(V^S_{j,t+1})$. Hence, the greater the welfare loss from infections, the greater their mobility response. If values of type S and type I are the same, then we have $\partial \ln(m^S_{ij,t})/\partial \alpha_{j,t+1} = 0$: people do not respond to infection risk.

In addition, the magnitude of the response is proportional to $1 - m^S_{ij,t}$, the total mobility rate to other destinations: the higher $1 - m^S_{ij,t}$ is, the greater the response of type S to an increase in infection rate in region j. Intuitively, a higher total mobility rate to other regions implies that region j becomes less attractive as a destination. Then, if the conditions in region j deteriorate, individuals respond more aggressively to such changes. Effectively, mobilities to other regions shield agents against the rising infection risk in region j.

3.4 Local and Global R_0

We next study the basic reproduction number, R_0, which determines if an emerging epidemic can be contained or not and characterizes disease dynamics at its outbreak. R_0 measures the expected number of new infections generated per infected individual near the DFE (Diekmann et al., 1990). Without cross-region mobility or interactions, the *local* R_0 of region i is:

$$R^L_{0,i} = \frac{\chi_i}{\gamma_i}, \tag{15}$$

where χ_i is the transmission rate capturing expected new infections per infected individual per period, and $1/\gamma_i$ is the expected infection duration.[13] Therefore, χ_i/γ_i is the expected new infections per infected individual over the course of infection. If $R^L_{0,i} > 1$, the epidemic continues spreading in region i, while $R^L_{0,i} < 1$ implies it will die out.

We generalize the notion of R_0 to an environment with human mobility across regions, following studies of multi-group SIRD models in which individuals may differ by age, gender, or other characteristics (Allen and Van den Driessche, 2008) or live in multiple regions (Antràs et al., 2023). We first combine equations (10) and (13) and rewrite the law of motion for type I as:[14]

$$I_{i,t} = \chi_i \frac{\overline{S}_{i,t}}{\overline{L}_{i,t}} \overline{I}_{i,t} + \left(1 - \gamma_i^R - \gamma_i^D\right) \overline{I}_{i,t} = \left(\chi_i \frac{\overline{S}_{i,t}}{\overline{L}_{i,t}} + 1 - \gamma_i^R - \gamma_i^D\right) \sum_{j=1}^{N} m^I_{ji} I_{j,t-1}. \tag{16}$$

Following Allen and Van den Driessche (2008), we vectorize the equation above near the DFE (hence $\frac{\overline{S}_{i,t}}{\overline{L}_{i,t}} = 1, \ \forall i$) and obtain

$$\mathbf{I}_{t+1} = \mathbf{F}\mathbf{I}_t + \mathbf{V}\mathbf{I}_t, \tag{17}$$

where $\mathbf{I_t} = [I_{1,t}, ..., I_{N,t}]'$, and $\mathbf{F} = diag(\chi_1, \chi_2, ..., \chi_N)\mathbf{m}^I$ and $\mathbf{V} = diag(1-\gamma_1, 1-\gamma_2, ..., 1-\gamma_N)\mathbf{m}^I$ while $\mathbf{m}^I = \{m^I_{ji}\}_{N\times N}$ is the mobility rate matrix of type I, and $\gamma_i = \gamma_i^D + \gamma_i^R, \ (1 \leq i \leq N)$, is the *removal rate* of type I in region i.

Intuitively, $\mathbf{F}$ is the transmission matrix, which captures new infections generated by transmission. $\mathbf{V}$ is the transition matrix, which captures the transition of type I to type R or D. In our model, individuals can move from one region to another, potentially generating new infections. This is captured by the mobility matrix $\mathbf{m}^I$, which appears in both $\mathbf{F}$ and $\mathbf{V}$. We follow Allen and Van den Driessche (2008) and define the *next-generation matrix* as

$$\mathbf{M} = \mathbf{F}\left(\mathbb{I} - \mathbf{V}\right)^{-1}, \tag{18}$$

where $\mathbb{I}$ is an identity matrix. This concept is borrowed from the literature on matrix population models that examine demographic dynamics and view infection as consecutive generations of infected people (Caswell, 2000). $\mathbf{M}$ connects the number of newly infected cases in different regions for consecutive periods. The basic reproduction number is given by the spectral radius of $\mathbf{M}$, the largest norm of its eigenvalues:

$$R_0^G \equiv \rho(\mathbf{M}) \equiv \max_{1\leq i\leq N} \{|\Lambda_i|\}, \tag{19}$$

where Λ_i is the eigenvalue of $\mathbf{M}$. We call R_0^G is the *global* R_0 of our multi-region economy.

[13] In this subsection on R_0, we follow the literature to assume the transmission rate is time-invariant.

[14] Note that we have used $m^I_{ij,t} = m^I_{ij}$ from Proposition 3,

The interpretation R_0^G has similarity with R_0^L in the tradition of SIRD models. First, the element on the i^{th} row and j^{th} column of $\mathbf{F}$, $\chi_i m_{ji}^I$, is the expected number of new cases in region i generated by an infected individual from region j. As for $\mathbf{V}$, the element on the i^{th} row and j^{th} column, $(1-\gamma_i)m_{ji}^I$, is the survival probability in region i for a type I individual originated from j. Then, the element on the i^{th} row and j^{th} column of $(\mathbb{I} - \boldsymbol{V})^{-1}$ is the expected duration of infection for an infected individual in region i who has moved from region j.[15] Therefore, the element on the i^{th} row and j^{th} column of $\mathbf{M}$ represents the expected number of new infections in region i generated by an infected individual originated from region j.

We next explain why R_0^G determines whether the pandemic can be contained or not in this multi-region environment. We consider a small outbreak across regions near the DFE, with the initial stock of infected individuals given by $\mathbf{I_0} = \delta_\mathbf{0}$. We call it the 0-th generation of infection. Then, the *expected* number of new cases directly generated by the 0-th generation, that is the first generation of cases is $\mathbf{I_1} = \mathbf{M}\delta_\mathbf{0}$. In general, the k-th generation of infection is $\mathbf{I_k} = \mathbf{M}\mathbf{I_{k-1}} = \mathbf{M}^k\delta_\mathbf{0}$. Therefore, the growth rate of $\mathbf{I_k}$ is controlled by the largest norm of eigenvalues of $\mathbf{M}$, namely R_0^G.[16] We next discuss the properties of R_0^G.

Proposition 5. *R_0^G falls between the minimum and maximum of local reproduction numbers, i.e.,* $\min_i R_{0,i}^L \leq R_0^G \leq \max_i R_{0,i}^L$.

Proof. See appendix A.1.6. □

This result implies that cross-region mobility generates epidemiological externalities by allowing regions to share infection risks and medical resources. People tend to move away from regions with high infection rates to regions with low infection rates, and from regions with low recovery rates to regions with high recovery rates, both tend to lower the infection globally. Intuitively, if everyone stays in the region with the lowest R_0^L, we have $R_0^G = \min_i R_{0,i}^L$. Similarly, if everyone stays in the region with the highest local R_0, we have $R_0^G = \max_i R_{0,i}^L$. In general, R_0^G falls between the maximum and minimum values of local R_0s when people can move across regions.[17]

Proposition 5 has important policy implications. First, regardless of the mobility frictions, if the largest local R_0 is less than 1, global R_0 is less than 1. Therefore, if the infectious disease can be controlled in the region with the worst situation, allowing free cross-region mobility will not cause a pandemic. It also provides rationales for prioritizing disease control in that region. If the pandemic can be controlled there, it can be contained globally. But this result does not imply we should give up disease control in other regions, which we investigate in the next proposition.

[15] Let ξ_i be the expected duration of infection in region i and $\xi = [\xi_1, ..., \xi_N]$, then ξ satisfies a recursive relationship $\xi = \iota + \xi V$, and $\iota = [1, ..., 1]_{1\times N}$. Immediately, we have $\xi = \iota\left(\mathbb{I} - \boldsymbol{V}\right)^{-1}$.

[16] We note that $\lim_{k\to\infty} \frac{\boldsymbol{I_k}}{\zeta^k} = \lim_{k\to\infty}(\frac{\boldsymbol{M}}{\zeta})^k I_0$ is finite if $\zeta \geq \rho(\boldsymbol{M})$.

[17] Because of idiosyncratic preference shocks, people will not all flow to the region with the lowest local R_0 or out of the region with the highest local R_0.

Proposition 6. *With a positive mobility matrix (i.e., $m^I_{ij} > 0,\ \forall\ i\&j$), a decrease in the transmission rate χ_i or an increase in the removal rate γ_i in any region decreases global R_0^G .*

Proof. See appendix A.1.7. □

Proposition 6 suggests that in a globalized world with cross-region mobility all regions matter, and improving infectious disease control in any region contributes to global disease control.

4 Quantitative Analysis

Our analyses of R_0 focus on the initial phase of the pandemic. To capture whole disease dynamics, we have to solve the sequential competitive equilibrium. For that purpose, this section develops the *Normalized Hat Algebra*. It enables us to solve the transitional path of the model and analyze disease and mobility dynamics.

4.1 Normalized Hat Algebra

Hat Algebra is a set of comparative static and counterfactual analysis methodologies. It was first proposed by Jones (1965) to solve trade models by log linearization. Instead of log linearization, Dekle et al. (2007) solved a trade model in terms of exact changes from the current equilibrium and called their method "Exact Hat Algebra" (EHA). However, EHA can only be applied to static or steady-state models. Caliendo et al. (2019) extended EHA and developed the method of "Dynamic Hat Algebra" (DHA) to solve models with intrinsic migration dynamics. Their idea is to transform the model in terms of relative time differences. They show that the DHA allows them to derive the transitional path and conduct counterfactual analyses without knowing all model fundamentals.

Our model does not belong to the class of models solvable by DHA. To see that, we follow Caliendo et al. (2019) to write type S's value function (equation 8) in terms of time differences:[18]

$$\dot{V}^S_{i,t+1} = \sum_{k=1}^{N} m^S_{ik,t} (\dot{V}^S_{k,t+2})^{\beta(1-\alpha_{k,t+2})} (\dot{V}^I_{k,t+2})^{\beta\alpha_{k,t+2}} \left(\frac{V^S_{k,t+1}}{V^I_{k,t+1}} \right)^{\beta(\alpha_{k,t+1}-\alpha_{k,t+2})} \dot{\mu}^{-1}_{ik,t+1}, \tag{20}$$

where $\dot{X}_{i,t+1} \equiv \frac{X_{i,t+1}}{X_{i,t}}$ is the proportional change of variable X from period t to $t+1$. However, the last term in the equation, $(V^S_{k,t+1}/V^I_{k,t+1})^{\beta(\alpha_{k,t+1}-\alpha_{k,t+2})}$, is still expressed in terms of the *level* of the value function and shows up whenever the probability of infection varies over time: $\alpha_{k,t+1} \neq \alpha_{k,t+2}$. Since $\alpha_{k,t+1}$ depends on the time-varying density of infected and susceptible individuals in each region, as specified in equation (12), the model cannot be rewritten in terms of

[18] See appendix B.4 for details of the derivation.

time differences. Only when $\alpha_{k,t+1} = \alpha_{k,t+2}$ for any t and k, that is the probability of infection is time-invariant, our model would fall into the class that can be solved by the DHA.

We have developed a new method which we call the Normalized Hat Algebra (NHA) to solve the model. The idea is to normalize the value functions by their respective steady-state values. For a variable x_t, we denote its normalized value as $\widehat{x}_t = x_t/x_{ss}$, while x_{ss} is its steady-state value. Using this normalization, we prove the following result.

Proposition 7. *Given the initial distribution of agents, G_0, and the steady state of the economy, the solution to the sequential competitive equilibrium solves equations (10), (11), (12), and (13), together with the following equations for all regions in all periods*

$$\begin{aligned}
\widehat{V}^S_{i,t} &= \sum_{j=1}^{N} m^S_{ij,ss} \left(\widehat{V}^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})} \left(\widehat{V}^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}} \left(\frac{V^I_{j,ss}}{V^S_{j,ss}}\right)^{\beta\alpha_{j,t+1}} \left(\widehat{\mu}_{ij,t}\right)^{-1}, \\
\widehat{V}^I_{i,t} &= \sum_{k=1}^{N} m^I_{ik,ss} \left(\widehat{V}^I_{k,t+1}\right)^{\beta(1-\gamma^R_k-\gamma^D_k)} \left(\widehat{V}^R_{k,t+1}\right)^{\beta\gamma^R_k} \widehat{\mu}^{-1}_{ik,t}, \\
\widehat{V}^R_{i,t} &= \sum_{k=1}^{N} m^R_{ik,ss} \left(\widehat{V}^R_{i,t+1}\right)^{\beta} \widehat{\mu}^{-1}_{ik,t},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{m}^S_{ij,t} &= \frac{\left(\widehat{V}^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})} \left(\widehat{V}^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}} \left(\frac{V^I_{j,ss}}{V^S_{j,ss}}\right)^{\beta\alpha_{j,t+1}} \left(\widehat{\mu}_{ij,t}\right)^{-1}}{\sum_{k=1}^{N} m^S_{ik,ss} \left(\widehat{V}^S_{k,t+1}\right)^{\beta\left(1-\alpha_{k,t+1}\right)} \left(\widehat{V}^I_{k,t+1}\right)^{\beta\alpha_{k,t+1}} \left(\frac{V^I_{k,ss}}{V^S_{k,ss}}\right)^{\beta\alpha_{j,t+1}} \left(\widehat{\mu}_{ik,t}\right)^{-1}}, \\
\widehat{m}^I_{ij,t} &= \frac{\left(\widehat{V}^I_{i,t+1}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)} \left(\widehat{V}^R_{j,t+1}\right)^{\beta\gamma^R_j} \left(\widehat{\mu}_{ij,t}\right)^{-1}}{\sum_{k=1}^{N} m^I_{ik,ss} \left(\widehat{V}^I_{k,t+1}\right)^{\beta(1-\gamma^R_k-\gamma^D_k)} \left(\widehat{V}^R_{k,t+1}\right)^{\beta\gamma^R_k} \widehat{\mu}^{-1}_{ik,t}}, \\
\widehat{m}^R_{ij,t} &= \frac{\left(\widehat{V}^R_{j,t+1}\right)^{\beta} \widehat{\mu}^{-1}_{ij,t}}{\sum_{k=1}^{N} m^R_{ik,ss} \left(\widehat{V}^R_{i,t+1}\right)^{\beta} \widehat{\mu}^{-1}_{ik,t}},
\end{aligned}$$

where $m^g_{ij,ss}$, $g \in \{S, I, R\}$ is the steady-state mobility rate of type g from region i to region j.

Proof. See appendix A.1.8 □

We note that $\widehat{V}^S_{i,t}$ contains $(V^I_{j,ss}/V^S_{j,ss})^{\beta\alpha_{j,t+1}}$, which is time-varying and not a "hat" variable. But the problem is simpler than in equation (20) using the DHA method: $V^I_{j,ss}/V^S_{j,ss}$, the steady-state ratio of value functions between type I and S, is time-invariant. Given the steady-state

mobility costs and instantaneous utilities, it can be obtained from equation (8) by replacing $\alpha_{j,ss} = 0\ (\forall j)$. Appendix B.5 presents the NHA algorithm in greater detail.

4.2 Model Dynamics

To investigate disease, demographic, and mobility dynamics, we employ the NHA method to simulate a symmetric 3-region economy: regions 1, 2, and 3, with the outbreak starting in Region 1. This setup provides a parsimonious structure to analyze mobility flows between regions with and without the disease outbreak, and between regions without the outbreak. The model parameters for this 3-region economy are provided in appendix Table B.3.

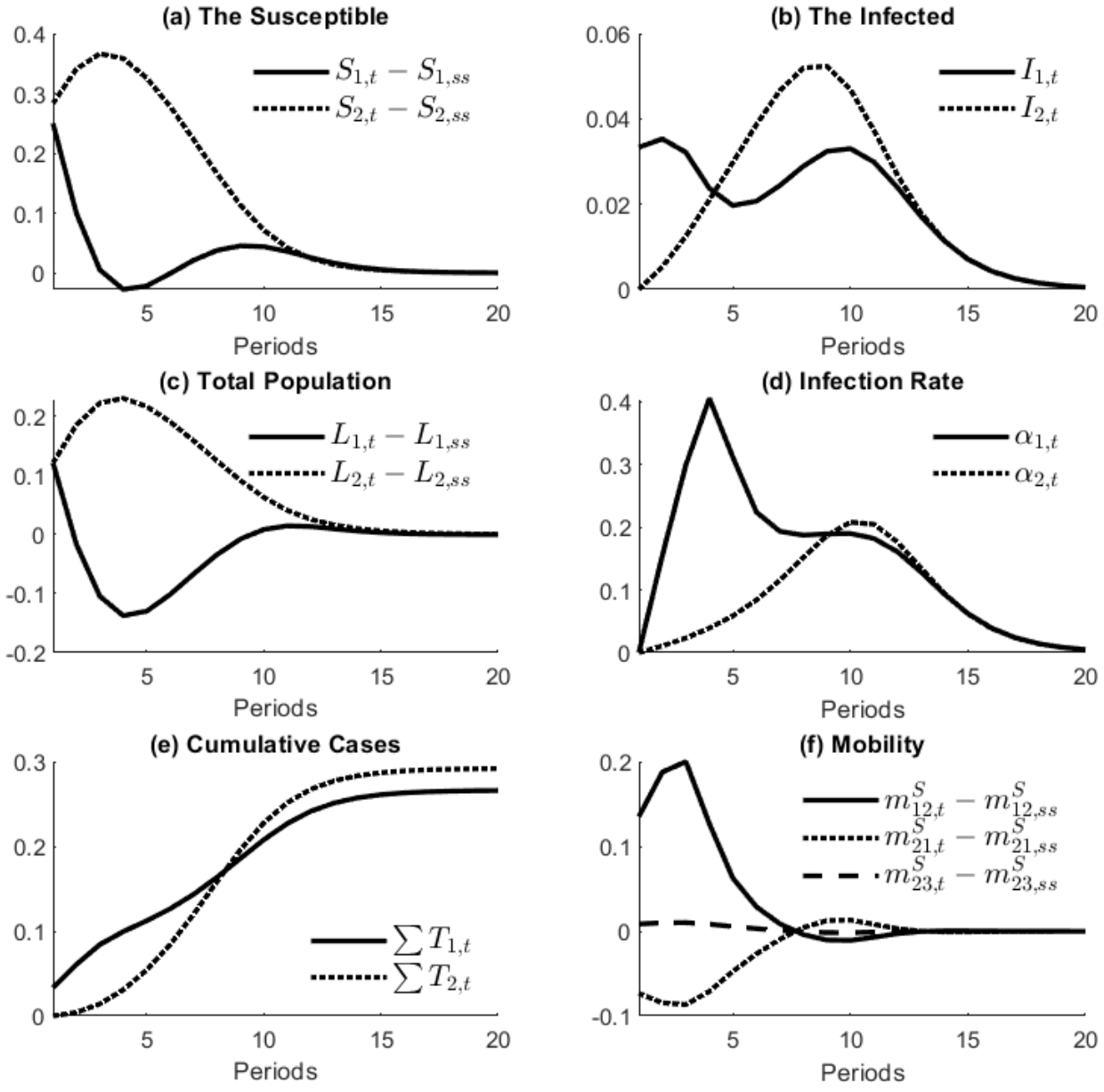
The simulation results are presented in Figure 5. We assume that 1% of the local population in Region 1 are initially infected.[19] In line with Fact 3 and Proposition 4, susceptible individuals rapidly move out of Region 1, which has the highest infection rate, to other lower-risk regions. In the other two regions, susceptible individuals avoid moving to Region 1 and prioritize regions without outbreaks (panel f). These changes in mobility lead to a decrease in the population and stock of susceptible individuals in Region 1 and a rise in the population and stock of susceptible individuals in other regions (panels a and c). Consequently, the density of infected individuals rises in Region 1, leading to an increase in the local infection rate. As infected individuals move to Regions 2 and 3, their respective local infection rates also rise (panel d). Over time, infected individuals either recover or pass away, resulting in a decline in their numbers in all regions (panel b). Despite the growing cumulative number of cases (panel e), infection rates eventually decreased in all regions. As the infection rate falls, susceptible individuals return to Region 1, and the economy converges to a new steady state.

A few interesting observations emerge from the simulation. First, there can be *multiple waves* of infections.[20] As shown in panel b, there are two peaks of infections in Region 1. As the region with the initial outbreak, infection quickly rises, which constitutes the first wave. But when infections gradually increase in the other two regions, the susceptible individuals move back to Region 1 (panels a and f), which pushes up infections in Region 1 and causes the second wave. Second, there are *overshootings* and *undershootings* in cross-region human mobilities. In panel f, we find that the initial outflows from Region 1 to other regions rise above their steady-state values while the inflows fall below their steady-state values. Finally, the region with the outbreak does not necessarily have the highest cumulative cases. At the steady state, the cumulative number of infected cases in Region 2 is higher than in Region 1 (panel e).

[19] Regions 2 and 3 have symmetric results. Results for Region 3 are omitted in the figure.

[20] Typically, SIR models can generate multiple waves of infection if immunity is temporary. If immunity is permanent, such models predict a single wave (Kissler et al., 2020). Despite assuming permanent immunity, we can generate multiple waves in a multi-region model just like Antràs et al. (2023).

Figure 5: Epidemic and Mobility Dynamics with an Outbreak in Region 1



Notes: These figures plot the simulation results from a three-region economy with an initial outbreak in Region 1 (1% of the local population infected). The model parameters are specified in appendix Table B.3. Figure (a) plots the stock of type S individuals relative to its steady state value; (b) plots the number of type I individuals; (c) plots the size of the local population relative to its steady state value; (d) plots the infection rate; (e) plots the cumulative number of cases in Regions 1 and 2 by summing the number of new cases across all periods; and (f) plots type S's mobility rate in and out of Region 1 and between the other 2 regions relative to the steady state.

5 Quantifying the COVID-19 Pandemic in the US

Our previous quantification focuses on a hypothetical three-region economy. This section extends the analyses by fitting our spatial epidemic model to the US economy of 50 states and Washington, DC at a bi-weekly frequency. We estimate the pre-pandemic economy using data from January 2020 and the pandemic economy using data from April to December 2020.[21]

In the following, we present strategies to estimate the key model parameters and the estimation results. We then use the estimated model to conduct counterfactual simulations and evaluate the effect of COVID-19 containment policies and cross-state mobility policies imposed by US governments on health and welfare outcomes. We find local containment policies flatten the US epidemic curve effectively but mobility control policies play a complex role: they reduce infection but have led to more deaths and lower social welfare.

5.1 Parameter Estimation

For some of the model parameters, we take them directly from the literature or calibrate them to match aggregate data moments. These include the discount rate, mobility elasticity, infection rate, and recovery rate. We develop strategies to estimate mobility costs and transmission rates and their elasticities with respect to containment policies, which are crucial for counterfactual policy evaluations. Finally, we also estimate the value of the deceased, which is necessary to solve the model. In the following, we will use $\widetilde{a}$ to denote the estimate of a parameter a.

Inter-temporal Discount Rate and Mobility Elasticity

We set the biweekly inter-temporal discount rate $\widetilde{\beta} = 0.9985$ to match a 4% annual interest rate. As far as we know, there is no estimation of mobility elasticity at the bi-week frequency.[22] Using US data, Caliendo et al. (2019) estimated a quarterly mobility elasticity of 5.34, a five-month mobility elasticity of 3.95, and an annual elasticity of 2.02. The mobility elasticity appears to rise with mobility frequency. We set the bi-week mobility elasticity at $\widetilde{\kappa} = 7.4$ by interpolating from the mobility elasticity estimates of Caliendo et al. (2019) to the bi-week level.

[21] This time window was chosen for three reasons. First, quality recovery data were unavailable until April 2020. Second, the US government started to roll out vaccination in December 2020 and more variants of the coronavirus started to emerge in late 2020. We evaluate the effect of vaccination in the US in Chen et al. (2022). A model extension with vaccination is presented in appendix B.9.

[22] Although we get access to mobility data at a bi-week frequency, we lack economic data at the same or similar frequencies that would have enabled us to estimate the elasticity.

Instantaneous Utilities and Pandemic-free Mobility Costs

We next estimate the instantaneous utility of each US state (i.e., $\{u_i\}_{i=1}^{51}$) and inter-state pandemic-free mobility costs (i.e., $\{\mu_{ij}\}_{i=1,j=1}^{51,51}$). They jointly determine inter-state mobility before the outbreak of the pandemic, given the inter-temporal discount rate and mobility elasticity. We estimate them jointly using *PlaceIQ* data from the first bi-week of January 2020, before the first COVID-19 case was reported in the US. We treat the observed mobility flows at that time as the mobility pattern when the US is pandemic-free.[23]

To estimate these two sets of parameters, we consider a version of our baseline model without pandemics. Essentially, individuals draw *i.i.d.* preference shocks following the same distribution as equation (5) in each period and move to the region that yields the highest expected utility in the following period, net of the mobility cost. The model details are as outlined in appendix B.6. At the steady-state equilibrium, the mobility rate from state i to state j, m_{ij}, relative to the probability of staying in state i, m_{ii}, is given by

$$\ln\left(\frac{m_{ij}}{m_{ii}}\right) = \beta \ln V_j - \beta \ln V_i - \ln \mu_{ij}, \tag{21}$$

where V_i is the value of staying in state i and μ_{ij} is the pandemic-free mobility cost of moving from state i to state j. We project the mobility costs on gravity variables, including bilateral distance, $dist_{ij}$, and an indicator of contiguity $cont_{ij}$ between two states, and a destination fixed effect:[24]

$$\ln \mu_{ij} = \alpha_0 + \alpha_1 \ln dist_{ij} + \alpha_2 cont_{ij} + \delta_j. \tag{22}$$

Let's denote $S_i = \beta \ln V_i$, equations (21) and (22) imply the following estimation equation:

$$\ln\left(\frac{m_{ij}}{m_{ii}}\right) = -\alpha_0 - \alpha_1 \ln dist_{ij} - \alpha_2 cont_{ij} + S_j - S_i - \delta_j + \zeta_{ij}, \tag{23}$$

where ζ_{ij} is the error term. Then, if we have estimated $\{\widetilde{S}_i,\ \widetilde{\alpha}_0,\ \widetilde{\alpha}_1,\ \widetilde{\alpha}_2,\ \widetilde{\delta}_j\}$ using the above equation, bilateral mobility costs and state-specific instantaneous utilities are given by:

$$\widetilde{\mu}_{ij} = \exp\left\{\widetilde{\alpha}_0 + \widetilde{\alpha}_1 \ln dist_{ij} + \widetilde{\alpha}_2 cont_{ij} + \widetilde{\delta}_j\right\} \tag{24}$$

$$\widetilde{u}_i = \kappa[\widetilde{S}_i/\beta - \ln \sum\nolimits_{j=1}^{51} \left(exp(\widetilde{S}_j)/\widetilde{\mu}_{ij}\right)]. \tag{25}$$

However, the implied mobility rates $\{\widetilde{m}_{ij}\}$ is not necessarily consistent with the observed pop-

[23] See appendix B.1 for detailed descriptions of how we construct the mobility matrix using *PlaceIQ* data.

[24] This specification captures asymmetric mobility costs between states. Similar specifications have been adopted in the estimation of mobility costs by Artuç et al. (2010), and trade costs by Waugh (2010) and Tombe and Zhu (2019).

ulation distribution across states, $\{L_i\}_{i=1}^{51}$, such that[25]

$$\sum_{i=1}^{51} \widetilde{m}_{ij} L_i = L_j. \tag{26}$$

Inspired by Anderson and van Wincoop (2003), who estimate the gravity equation for trade flows together with a market-clearing constraint, we adopt an iterative method to ensure the estimates are consistent with the add-up constraint (26). In particular, we initialize $\{\widetilde{S}_i^0\}$ and take $\{\widetilde{\alpha}_0, \widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{\delta}_j\}$ from the OLS estimate of equation (23). We then estimate mobility costs by equation (24) and update $\{\widetilde{S}_i\}$ by solving equation (26), with the mobility rates given by:

$$\widetilde{m}_{ij} = \frac{exp(\widetilde{S}_j)/\widetilde{\mu}_{ij}}{\sum_{k=1}^{51} exp(\widetilde{S}_k)/\widetilde{\mu}_{ik}}. \tag{27}$$

The iteration stops when differences in the estimated $\{\widetilde{S}_i\}$ between two steps are sufficiently small.

We get an estimated distance elasticity of $\widetilde{\alpha}_1 = 1.02$, which implies that a one-percentage increase in inter-state distance raises the mobility cost by 1.02%. The contiguity dummy has an estimated coefficient of $\widetilde{\alpha}_2 = -1.09$, which implies that sharing a common border reduces cross-state mobility costs by 66.4% on average. Appendix Figure B.1 plots the distribution of the estimated bilateral mobility costs, which has a roughly bell-shaped distribution. Appendix Figure B.2 validates the estimated value of each state, V_i, by plotting it against the state average real income (in logs). The two are positively correlated, with a correlation of 0.55.

Mortality and Recovery Rates of COVID-19

We calculate state-specific COVID-19 mortality rates $\{\widetilde{\gamma}_i^D\}$ by taking the ratio of total deaths over total infected cases by December 2020. It averages at 1.64% nationally and ranges from 0.45% in Alaska to 3.65% in New Jersey. Recovery rates $\{\widetilde{\gamma}_i^R\}$ are estimated similarly. We take the ratio of total recovered COVID-19 cases over the total number of infected cases during the sample period for each state. The average estimated recovery rate is 69.7% for 30 states with available recovery data. We set the recovery rates for those states with missing recovery data at this average.[26]

[25]This is probably due to temporal shocks to the economy such that the observed mobility pattern deviates from the steady-state equilibrium. If we adopted these estimates, the implied population distribution could be degenerated such that certain states would have a population close to zero.

[26]These numbers are close to Atkeson (2020). He noticed that it takes an average of 18 days for the infected to recover or die. It implies a removal rate $\gamma_i^R + \gamma_i^D = \frac{14}{18} \approx 77.8\%$. In our case, the average removal rate is $1.64\% + 69.7\% = 71.34\%$.

Transmission Rates, Mobility Costs and Containment Policies of COVID-19

We categorize the containment policies of COVID-19 into two types: local containment aimed at reducing within-region infection risk (e.g., public place closures), and cross-region mobility policies that restrict inflows from other states.

We first identify the effect of local containment policies on transmission rates, i.e., $\chi_{i,t}$, the probability of infection upon contact between infected and susceptible individuals. In our model, $\chi_{i,t}$ appears in equation (11), the matching function between susceptible and infected individuals. We parameterize it as

$$\ln \chi_{i,t} = \theta \cdot Policy_{i,t}^{local} + \rho_i + \rho_t, \tag{28}$$

which varies with the local containment policy index $Policy_{i,t}^{local}$, state fixed effects ρ_i, and time fixed effects ρ_t. Substituting equation (11) into (12) using (28), we can obtain the following equation for estimation:

$$\ln \alpha_{i,t} (\frac{\bar{I}_{i,t}}{\bar{L}_{i,t}})^{-1} = \theta \cdot Policy_{i,t}^{local} + \rho_i + \rho_t + \varepsilon_{i,t}, \tag{29}$$

where $\alpha_{i,t}$ is the infection rate, and $\frac{\bar{I}_{i,t}}{\bar{L}_{i,t}}$ is density of infected individuals. Conditional on $\frac{\bar{I}_{i,t}}{\bar{L}_{i,t}}$, equation (29) relates $\alpha_{i,t}$ to local containment policies, state fixed factors such as infrastructure, and time fixed effects capturing things like seasonality and economic cycles (Adda, 2016). We estimate it using state-level epidemiological and policy data and obtain an estimate of $\widetilde{\theta} = -0.29\%$. Therefore, increasing the local containment policy by one point would reduce the transmission rate by about 0.30%. Using the estimated $\widetilde{\theta}$, we infer the transmission rate $\{\widetilde{\chi}_{i,t}\}$ using equation (28).

In the same spirit, we parameterize mobility costs during the pandemic as follows:

$$\ln \mu_{ij,t} = \beta_0 + \beta_1 \ln dist_{ij} + \beta_2 cont_{ij} + \delta_j + \eta \cdot Policy_{j,t}^{mob}. \tag{30}$$

In addition to the gravity variables in equation (22), we include a time-variant policy variable $Policy_{j,t}^{mob}$, which captures the effect of mobility restrictions on mobility costs during the pandemic. From equation (9), we find that mobility rates during the pandemic satisfy:

$$\ln \frac{m_{ij,t}^g}{m_{ii,t}^g} + \ln \frac{m_{ji,t}^g}{m_{jj,t}^g} = -\ln \mu_{ij,t} - \ln \mu_{ji,t}, \tag{31}$$

for type $g \in \{S, I, R\}$ agent. Assuming that the majority of the population was susceptible during our sample period,[27] we aggregate mobility across all types of agents and net out the pre-pandemic mobility in equation (21). As detailed in appendix B.7, we find that aggregate mobility approxi-

[27] Given that we examine the early phase of the COVID-19 pandemic, with less than 6% of the US population infected at the end of our sample (December 2020), the approximation is quite accurate.

mately satisfies the following equation

$$\ln \frac{m_{ij,t}}{m_{ii,t}} - \ln \frac{m_{ij}}{m_{ii}} + \ln \frac{m_{ji,t}}{m_{jj,t}} - \ln \frac{m_{ji}}{m_{jj}} \approx -\eta \cdot (Policy_{i,t}^{mob} + Policy_{j,t}^{mob}). \tag{32}$$

Intuitively, differences in the mobility rates between any two states relative to the probability of staying in the respective local state before and during the pandemic are driven by policies restricting cross-state mobility. We estimate the above equation by combining mobility data with policy data on mobility control. We obtain an estimate of $\widetilde{\eta} = 1.56\%$. It implies that leveling up the mobility control policy by one point increases mobility costs by around 1.60%. Using the estimated $\widetilde{\eta}$, we can then infer mobility costs $\{\widetilde{\mu}_{ij,t}\}$ using the equation (30).

The Value of the Deceased

Using linear approximation, we show in appendix B.8 that the difference in the mobility rates of susceptible and recovered individuals from state i to j relative to their respective probability of staying in i is related to the value of the deceased:

$$\ln \frac{m_{ij,t}^S}{m_{ii,t}^S} - \ln \frac{m_{ij,t}^R}{m_{ii,t}^R} \approx a_{ij,t} + \ln V^D \cdot b_{ij,t}, \tag{33}$$

where $a_{ij,t}$ and $b_{ij,t}$ are functions of model parameters β, γ^R, γ^D, and model outcomes $m_{ij,t}^R$, $V_{j,t}^R$ and $\alpha_{j,t}$. The intuition is as follows. Since susceptible individuals can get infected and die from infection while recovered individuals have immunity, they respond differently to disease conditions at the destination when they move. In particular, the larger the value of the deceased (a higher V^D), the smaller the difference in their mobility rates. In the extreme scenarios that death brings no well-being loss (i.e., $V^D = V_{j,t+1}^R$) or the death rate is zero (i.e., $\gamma^D = 0$), our model implies $\ln \frac{m_{ij,t}^S}{m_{ii,t}^S} - \ln \frac{m_{ij,t}^R}{m_{ii,t}^R} = 0$. Therefore, without the death threat, there is no difference in the mobility pattern of the susceptible and recovered individuals.

To estimate equation (33), we first substitute the estimated instantaneous utilities $\{\widetilde{u}_i\}$ and mobility costs $\{\widetilde{\mu}_{ij,t}\}$ into equations (8) and (9) to infer the value function and mobility rates of recovered individuals, which are denoted as $\widetilde{V}_{j,t}^R$ and $\widetilde{m}_{ij,t}^R$, respectively. Together with the observed infection rate in each region (i.e., $\alpha_{j,t}$) and other model parameters and outcomes necessary to compute $a_{ij,t}$ and $b_{ij,t}$, we obtain their estimates, which are denoted as $\widetilde{a}_{ij,t}$ and $\widetilde{b}_{ij,t}$, respectively. Finally, we use the observed cross-state mobility rate (i.e., $m_{ij,t}$) as a proxy for $m_{ij,t}^S$, the mobility rate of susceptible individuals,[28] and obtain the following equation for estimation:

[28]This is consistent with the discussion in the previous footnote that the majority of the US population were susceptible during our sample period.

$$\ln \frac{m_{ij,t}}{m_{ii,t}} - \ln \frac{\widetilde{m}^R_{ij,t}}{\widetilde{m}^R_{ii,t}} - \widetilde{a}_{ij,t} = \omega \cdot \widetilde{b}_{ij,t} + \xi_{ij,t}, \tag{34}$$

while $\xi_{ij,t}$ is the error term, and $\omega = \ln V^D$ according to equation (33). Appendix Figure B.3 plots the left-hand side of equation (34) against $\widetilde{b}_{ij,t}$, which shows a very tight linear relationship between the two. Therefore, our linear approximation is quite accurate.

An OLS estimation of equation (34) yields an estimate of $\widetilde{\omega} = -7.32$. It implies that $\widetilde{V}^D = exp(-7.32) \approx 0.066\%$. Therefore, the value of the deceased is about 0.7% of the average welfare before the pandemic.[29]

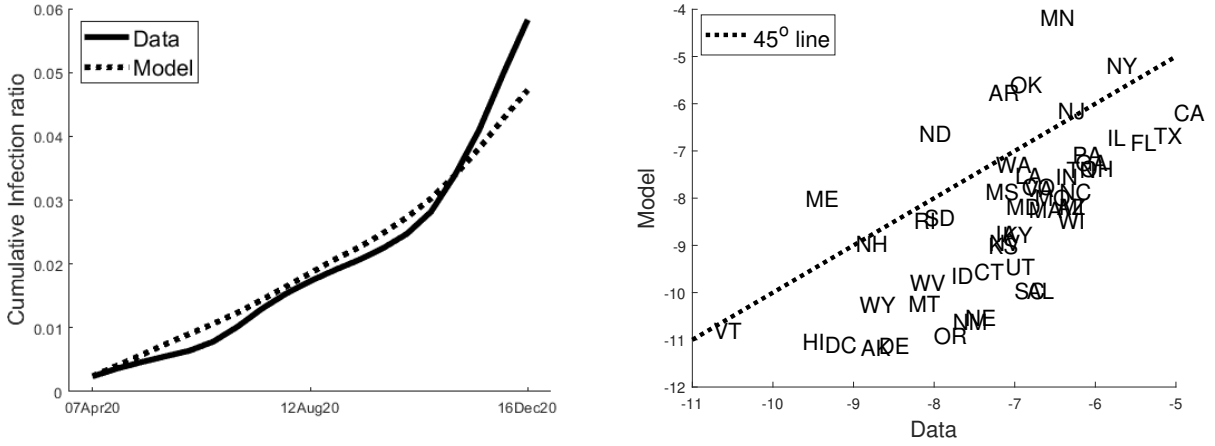
5.2 Counterfactual Analysis

In this subsection, we use the estimated quantitative model to evaluate the effect of COVID-19 containment policies on social welfare and health outcomes. We will focus on two types of policies: policies that restrict cross-state mobility, and intra-state containment policies that restrict local contacts between people.

Figure 6: Model Fit: Cumulative COVID-19 Infection Rate

(a) Cumulative Infection Rate of the US

(b) Cumulative Infection Rate by State



Notes: Figure (a) plots the cumulative share of the US population infected by COVID-19 in the data and as predicted by the model. Figure (b) plots the cumulative share of the US population infected by COVID-19 for each state (in logarithm), as predicted by the model against the data.

[29]Note that when we estimate the instantaneous utility of each state by fixed effects using estimating equation (23), we adopt the following normalization that $\sum_{i=1}^{N} L_i \ln(V_i) = 0$.

Validation

We first validate the estimated baseline quantitative economy by comparing the estimated numbers of infected cases with their actual numbers. We simulate the model from the first bi-week of April 2020 until it is sufficiently close to the steady state.[30] Figure 6 (a) plots the cumulative share of the US population infected by the coronavirus from April to December 2020 as generated by the model and observed in the data. Our model performs reasonably well in capturing the disease dynamics as the model-generated series tracks the data series closely. Our model also performs reasonably well in fitting the cumulative infection rates across US states. As shown in Figure 6 (b), states with higher cumulative infection rates are predicted to accumulate relatively more cases by the model.[31]

Counterfactuals

Baseline Economy Before examining the policy counterfactuals, we first look into the estimated baseline economy. The simulation results are presented in column (1) of Table 3. Our model predicts that around 82.1 million US residents will get infected eventually, leading to a cumulative death of 4.50 million.[32] To capture the disease dynamic in the US, the solid line in Figure 7 (b) plots the predicted epidemic curve for the first 100 bi-weeks, as represented by the number of newly infected cases relative to the US population. Similar to the three-region economy simulated in Figure 5 (b), the quantitative model predicts multiple waves of infections in the US, a feature that we also observe for COVID-19 in Figure 2 (a).[33]

Exogenous Mobility The first counterfactual scenario considers a model that removes endogenous human mobility by fixing the mobility rates at their pre-pandemic levels. Hence, agents do not respond to changes in infection rates or differential death or recovery rates across states and over time. Column (2) of Table 3 presents the simulated results of this scenario. Compared to the

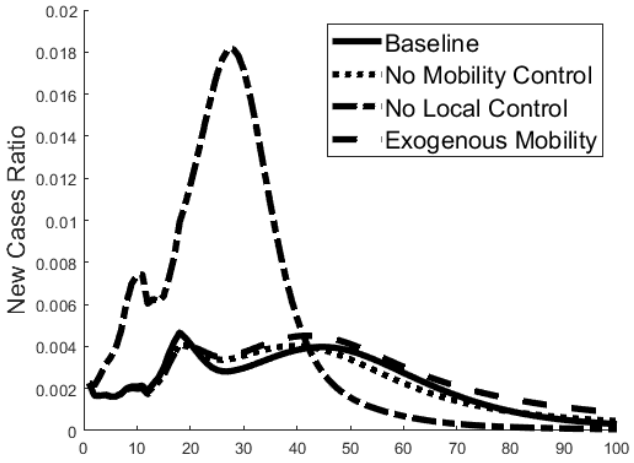
[30] We simulate the model 300 periods forward, which corresponds to 600 weeks or 11.5 years in reality.

[31] At the end of our sample, the model predicted cumulative national infection rate is about 5.0%, which is about 1.0% less than in the data. This is likely to be a result of the emergence of the more infectious variants of the coronavirus in late 2020, such as the *Alpha* variant.

[32] As of December 2023, the number of COVID-19 cases in the US was about 103 million according to Our World in Data and the number of deaths is about 1.16 million according to the US CDC COVID Data Tracker. Hence, we underestimate the number of cases but overestimate the number of deaths. There are a few reasons: (1) A few more infectious but less deadly variants of the coronavirus emerged since the end of our data sample period, including the Alpha, Delta, and Omicron variants, which pushed up infections (Carvalho et al., 2021); (2) Vaccination programs were introduced in late December 2020, which have reduced the number of infections and deaths (Chen et al., 2022); (3) Immunity weakens over time in reality and people can be reinfected (Chemaitelly et al., 2023). Our model assumes permanent immunity, which tends to reduce the number of infections and deaths.

[33] The model generates roughly four waves with the final wave peaking around the 50th periods, which is about two years since the start of the pandemic. This is less than the actual number of waves in reality. Again, this could be due to the rise of new variants and weakening immunity over time.

Figure 7: Epidemic Curves: Baseline and Counterfactuals



Notes: This figure displays the baseline and counterfactual epidemic curves of COVID-19, as represented by the number of newly infected cases divided by the US population in each bi-week. The first bi-week starts on 07 April 2020.

baseline in column (1), infected cases go up by another 31.3 million to 113.4 million, and deaths increase by another 1.28 million to 5.78 million. In other words, endogenous mobility reduces infections by 27.6% and deaths by 22.1%. In addition to more infections and deaths, Figure 7 (b) indicates that removing endogenous mobility prolongs the pandemic. Intuitively, if people could choose where to move, they would avoid moving to states with high infection and tend to move to states with low infection rates. This echoes the prior literature's finding on behavioral response as a source of social distancing (Farboodi et al., 2021; Bisin and Moro 2022; Antràs et al., 2023). However, this is not the end of the story: people also prefer moving to states with high recovery and low death rates than to states with low recovery and high death rates. Such "self-healing" mobility also helps reduce aggregate infections and deaths. Finally, due to the higher infection and death toll, the social welfare is about 18.4% below the baseline.[34]

No Cross-region Mobility Control In the second counterfactual scenario, we remove cross-region mobility control policies, i.e., setting $\eta = 0$ in equation (30), and keep other parameters the same as the baseline model. Given the estimated $\widetilde{\eta} = 1.56\%$, mobility costs are lower without mobility control policies. According to column (3) of Table 3, it leads to more infections, which increase by about 8.9 million, but lower deaths, which drop by 2.48 million as compared to the baseline economy in column (1). Intuitively, mobility control policies reduced contact between

[34]The social welfare is measured by the sum of value functions across all types of agents in all states at the steady state, weighted by the population.

the infected and susceptible individuals across states, which tends to reduce infection. However, they also create friction for infected individuals moving from states with low recovery rates to states with high recovery rates, and from states with high death rates to states with low death rates, denting the "self-healing" effect. Therefore, removing these mobility frictions tends to reduce deaths. Combined with the lower mobility costs, the US social welfare would go up by as much as 69.21% as compared to the baseline economy. Hence, the US would have been better off without the cross-state mobility control policies during the COVID-19 pandemic.

No Local Containment Policies In the final counterfactual scenario, we remove local containment policies, i.e., setting $\theta = 0$ in equation (28), and keep other parameters the same as the baseline model. Given the estimated $\widetilde{\theta} = -0.29\%$, transmission rates are higher without local containment policies, which leads to many more infections and deaths. According to column (4) of Table 3, infection amounts to 144.9 million, which is 62.8 million more than the baseline. Deaths go up to 5.81 million, which is 1.33 million more than the baseline. As evident from Figure 7 (b), infections rise sharply without the local containment policies, reaching a peak several folds higher than the peak in the baseline and the second counterfactual scenarios. Hence, local containment policies appear more effective than mobility control policies in "*flattening the epidemic curve*." The pandemic process is shorter without local containment policies, as the sharp outbreak quickly exhausts the susceptible individuals in the economy. However, due to the higher death toll, the welfare loss is 3.00% larger than the baseline result.

To sum up, endogenous mobility appears to generate a "self-containment" effect such that people voluntarily avoid moving to high-risk regions and a "self-healing" effect such that people prefer moving to regions with high recovery and low death rates. Mobility control policies, aiming to reduce infection by reshaping mobility incentives, reduce infections but create hurdles for "self-healing" mobility, which backfires on health outcomes and social welfare. Our estimation implies that removing cross-state mobility control policies would have led to slightly more infections but a lower number of deaths and higher social welfare. In contrast, without local containment policies, there would have been more infections and deaths, leading to lower social welfare.

Robustness Checks

To check the robustness of the result discussed above, we conduct a few simulations using alternative values for model parameters. The results are presented in appendix Table B.4. In row (a), we reduce the discount factor β from 0.9985 to 0.96. Consistent with the findings from Table 3, we find that without endogenous mobility, the number of cases and deaths go up, while removing cross-state mobility control policies raises infections but saves lives, and removing local containment policies raises both infections and deaths. In row (b), we replace the value of the deceased

Table 3: Counterfactual Simulations on the COVID-19 Pandemic in the United States

Scenarios	(1) *Baseline*	(2) Exogenous mobility	(3) No cross-state mobility control	(4) No local containment
Total infection (millions)	82.1	113.4	91.0	144.9
Total deaths (millions)	4.50	5.78	2.02	5.83
Welfare relative to *Baseline*		-18.37%	69.21%	-3.00%

Notes: This table shows the aggregate cumulative infections and deaths and social welfare loss in the United States until the COVID-19 pandemic vanishes in the long run, as predicted by our model. Column (1) presents simulation results using the baseline estimated parameters. Column (2) removes endogenous mobility by fixing the mobility rates at their pre-pandemic level, which was observed in January 2020. Column (3) removes cross-state mobility control policies from the baseline economy. Column (4) removes local containment policies from the baseline economy.

(in logarithm) from -7.32 to -10. While the pattern is consistent with Table 3, the predicted numbers of deaths get smaller with endogenous mobility as death has become more costly. Row (c) replaces the estimated state-specific death rates with a homogeneous value of 0.0164, and row (d) replaces the estimated state-specific recovery rates with a homogeneous value of 0.697. The welfare patterns remain the same as Table 3, but death goes up from 1.90 million to 2.16 million rather than falls when we remove cross-state mobility control in row (d). Hence, it is important to have spatial heterogeneity in recovery rates across states to generate the "self-healing" effect.

6 Discussion

Optimal Mobility Control Policies

So far, we have been focusing on evaluating the observed policies conducted by US state governments. A natural question is whether these policies have been optimal or not from the perspective of a benevolent federal government. Although the issue of optimal policies is not the focus of our analyses, we attempt to provide further insights about cross-state mobility policies using our model.[35] To simplify the analysis, we assume that the federal government maximizes social welfare at the steady state by varying mobility costs simultaneously across all regions, measured by μ, with mobility costs increasing with μ.

Appendix Figure B.4 shows the optimal mobility policies we have found for the three-region economy simulated in Section 4. Over a wide range of parameters, we find that there is no middle ground for optimal mobility control policy. The choice of policymakers can be characterized by a "*phase diagram*," such that the mobility friction μ is set as low as possible (the "Loose contain-

[35] See Fajgelbaum et al. (2021) study the optimal lockdown in commuting networks within cities, and Farboodi et al. (2021) for optimal social distancing policy over time.

ment" phase) in one parameter space, or as high as possible (the "Strict Containment" phase) in the other. Near the boundary of parameter combinations that policymakers are indifferent between the two phases, a small change in the parameters can trigger a *U-turn* in policies such that the policymaker switches from strict containment to loose containment on mobility.

We examine how the optimal mobility control policy varies with model parameters. We focus on the mortality rate (γ^D) and the transmission rate (χ) of infectious diseases, and the value of the decreased (V^D). In particular, as γ^D increases such that infectious diseases become more deadly, or V^D decreases such that the value of life rises, policymakers are more likely to adopt strict containment policies. Finally, as χ increases such that infectious diseases become more transmittable and difficult to contain, policymakers tend to prefer loose containment policies. In this case, we have to coexist with the infectious disease.

Model Extensions and Limitations

Our model is quite versatile. In appendix B.9, we discuss an extension incorporating vaccinations. To focus on mobility, we have removed inter-regional trade, which can be added following the standard Eaton and Kortum (2002) setup.[36] However, we do not have high-frequency inter-state trade data to estimate such a model. And we do not expect including inter-regional trade will change the model prediction on R_0. Because R_0 is determined by infected individuals' law of motion equation (16), which does not interact directly with inter-regional trade.

To simplify the analysis, we have assumed that the instantaneous utility of each state does not vary over time or change with disease conditions and containment policies. This assumption can be relaxed, for example, by assuming that the instantaneous utility depends on local containment policies. If local containment policies hurt the instantaneous utility, we could have overestimated the welfare loss of removing such policies. Similarly, we have assumed that infection does not reduce the instantaneous utility. Without taking such utility loss into account, we could have overestimated the welfare gains of removing cross-state mobility control and underestimated the welfare loss of removing local containment policies.

7 Conclusion

Rapid globalization and urbanization have generated enormous gains from human mobility, but may also have led to more frequent outbreaks of infectious diseases. We develop a theoretical framework to jointly study infectious diseases and human mobility. The model allows us to analyze how human mobility responds to infectious diseases, which in turn reshape human mobility. We

[36] See Redding and Rossi-Hansberg (2017) for a model with inter-regional trade and mobility.

characterize the theoretical properties of R_0. Unlike Antàs et al. (2023), the epidemiological externality is positive for regions with high local R_0s in our model. Endogenous cross-region mobility allows people to move to regions with low infection and death rates and high recovery rates, which serves as a risk and resource sharing mechanism between regions. The resulting global R_0 is between the minimum and maximum values of local R_0s. Due to the externality brought by cross-region mobility, improved disease control in any region reduces the global R_0.

We enrich the existing toolboxes of quantitative spatial models by developing the NHA method, which solves dynamic mobility models in which agent types change endogenously over time. We use it to study disease, demographic, and mobility dynamics. We find that the pandemic can have multi-waves and mobility flows can overshoot or undershoot their steady-state levels.

Applying the model to the US, we evaluate the health and welfare effect of cross-state mobility control and local containment policies via counterfactual simulations. We find that if restrictions on cross-state mobility were removed, the US would have fewer COVID-19 deaths and higher welfare, despite having slightly more infections. Mobility control policy might have reduced flows from regions with high infection rates, but it reduced mobility to states with low death rates and high recovery rates, impeding "self-healing" mobility. In addition, our simulation results indicate that there may be no middle ground for optimal mobility control policies. Hence, policymakers should be prudent in implementing mobility control policies during pandemics.

References

Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M.D., 2021. Optimal Targeted Lockdowns in a Multigroup SIR Model. *American Economic Review: Insights*, 3(4), pp.487-502.

Adda, J., 2016. Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data. *The Quarterly Journal of Economics*, 131(2), pp.891-941.

Allen, L.J., and Van den Driessche, P., 2008. The Basic Reproduction Number in Some Discrete-Time Epidemic Models. *Journal of Difference Equations and Applications*, 14(10-11), pp.1127-1147.

Allen, T., and Arkolakis, C., 2014. Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3), pp.1085-1140.

Alvarez, F.E., Argente, D., and Lippi, F., 2021. A Simple Planning Problem for COVID-19 Lockdown, Testing, and Tracing. *American Economic Review: Insights*, 3(3), pp.367-382.

Anderson, J.E. and Van Wincoop, E., 2003. Gravity with Gravitas: A Solution to The Border Puzzle. *American Economic Review*, 93(1), pp.170-192.

Anderson, S.P., De Palma, A., and Thisse, J.F., 1992. Discrete Choice Theory of Product Differentiation. MIT Press.

Antràs, Pol, Stephen J. Redding, and Esteban Rossi-Hansberg. 2023. Globalization and Pandemics. *American Economic Review*, 113 (4): 939-81.

Argente, D., Hsieh, C.T. and Lee, M., 2022. The Cost of Privacy: Welfare Effects of the Disclosure of Covid-19 Cases. *The Review of Economics and Statistics*, 104(1), pp.176-186.

Arellano, C., Bai, Y. and Mihalache, G.P., 2023. Deadly Debt Crises: COVID-19 in Emerging Markets. *The Review of Economic Studies*, rdad058, https://doi.org/10.1093/restud/rdad058

Artuç, E., Chaudhuri, S., and McLaren, J., 2010. Trade Shocks and Labor Adjustment: A Structural Empirical Approach. *American Economic Review*, 100(3), pp.1008-45.

Atkeson, A., 2020. What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios (No. w26867). *National Bureau of Economic Research*.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J. and Vespignani, A., 2009. Multiscale Mobility Networks and the Spatial Spreading of Infectious Diseases. *Proceedings of the National Academy of Sciences*, 106(51), pp.21484-21489.

Baldwin, R., 2020. It's not Exponential: An Economist's View of the Epidemiological Curve. *VoxEU.org, 12*.

Berger, D.W., Herkenhoff, K.F., and Mongey, S., 2020. An SEIR Infectious Disease Model with Testing and Conditional Quarantine (No. w26901). *National Bureau of Economic Research*.

Bisin, A., and Moro, A., 2022. JUE Insight: Learning Epidemiology by Doing: The Empirical Implications of a Spatial-SIR Model with Behavioral Responses. *Journal of Urban Economics*, 127, p.103368.

Caliendo, L., Dvorkin, M., and Parro, F., 2019. Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, 87(3), pp.741-835.

Caliendo, L., Opromolla, L.D., Parro, F., and Sforza, A., 2021. Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement. *Journal of Political Economy*, 129(12), pp.3491-3545.

Caswell, H., 2000. Matrix Population Models (Vol. 1). Sunderland, MA: Sinauer.

Chen, X., Huang, H., Ju, J., Sun, R. and Zhang, J., 2022. Impact of Vaccination on the COVID-19 Pandemic in US States. *Scientific Reports*, 12(1), p.1554.

Chemaitelly H, Ayoub HH, Tang P, Coyle P, Yassine HM, Al Thani AA, Al-Khatib HA, Hasan MR, Al-Kanaani Z, Al-Kuwari E, Jeremijenko A. Long-term COVID-19 Booster Effectiveness by Infection History and Clinical Vulnerability and Immune Imprinting: A Retrospective Population-Based Cohort Study. *The Lancet Infectious Diseases*. 2023 Mar 10.

Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A.P., Mu, K., Rossi, L., Sun, K., and Viboud, C., 2020. The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak. *Science*, 368(6489), pp.395-400.

Carvalho, T., Krammer, F. and Iwasaki, A., 2021. The First 12 Months of COVID-19: a Timeline of Immunological Insights. *Nature Reviews Immunology*, 21(4), pp.245-256.

Couture, V., Dingel, J.I., Green, A., Handbury, J. and Williams, K.R., 2022. JUE Insight: Measuring Movement and Social Contact with Smartphone Data: A Real-Time Application to COVID-19. *Journal of Urban Economics*, 127, p.103328.

Dekle, R., Eaton, J., and Kortum, S, 2007. Unbalanced Trade. *American Economic Review*, 97(2), pp.351-355.

Diekmann, O., Heesterbeek, J.A.P., and Metz, J.A., 1990. 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), pp.365-382.

Diekmann, O., Heesterbeek, J.A.P., and Roberts, M.G., 2010. The Construction of Next-Generation Matrices for Compartmental Epidemic Models. *Journal of the Royal Society Interface*, 7(47), pp.873-885.

Eaton, J. and Kortum, S., 2002. Technology, Geography, and Trade. *Econometrica*, 70(5), pp.1741-1779.

Eichenbaum, M.S., Rebelo, S., and Trabandt, M., 2021. The Macroeconomics of Epidemics. *The Review of Financial Studies*, 34(11), pp.5149-5187.

Eubank, S., Guclu, H., Anil Kumar, V.S., Marathe, M.V., Srinivasan, A., Toroczkai, Z. and Wang, N., 2004. Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature*, 429(6988), pp.180-184.

Fajgelbaum, P.D., Khandelwal, A., Kim, W., Mantovani, C., and Schaal, E., 2021. Optimal Lockdown in a Commuting Network. *American Economic Review*: Insights, 3(4), pp.503-22.

Fang, H., Wang, L., and Yang, Y., 2020. Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-ncov) in China. *Journal of Public Economics*, 191, p.104272.

Farboodi, M., Jarosch, G., and Shimer, R., 2021. Internal and External Effects of Social Distancing in a Pandemic. *Journal of Economic Theory*, 196, p.105293.

Giannone, E., Paixão, N. and Pang, X., 2022. JUE Insight: The Geography of Pandemic Containment. *Journal of Urban Economics*, 127, p.103373.

Glaeser, E.L., Gorback, C., and Redding, S.J., 2022. JUE Insight: How Much Does COVID-19 Increase with Mobility? Evidence from New York and Four Other US Cities. *Journal of Urban Economics*, p.103292.

Goolsbee, A., and Syverson, C., 2021. Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020. *Journal of Public Economics*, 193, p.104311.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., and Tatlow, H., 2021. A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4), pp.529-538.

Harko, T., Lobo, F.S., and Mak, M.K., 2014. Exact Analytical Solutions of the Susceptible-Infected-Recovered (SIR) Epidemic Model and of the SIR Model with Equal Death and Birth Rates. *Applied Mathematics and Computation*, 236, pp.184-194.

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L.Y., Hultgren, A., Krasovich, E., and Lau, P., 2020. The Effect of Large-Scale Anti-Contagion Policies on the COVID-19 Pandemic. *Nature*, 584(7820), pp.262-267.

Jones, R.W., 1965. The Structure of Simple General Equilibrium Models. *Journal of Political Economy*, 73(6), pp.557-572.

Kermack, W.O., and McKendrick, A.G., 1927. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London*. Series A, 115(772), pp.700-721.

Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H., and Lipsitch, M., 2020. Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period. *Science*, 368(6493), pp.860-868.

Li, C.K., and Schneider, H., 2002. Applications of Perron-Frobenius Theory to Population Dynamics. *Journal of Mathematical Biology*, 44(5), pp.450-462.

Kleinman, B., Liu, E. and Redding, S.J., 2023. Dynamic Spatial General Equilibrium. *Econometrica*, 91(2), pp.385-424.

Longini Jr, I.M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D.A. and Halloran, M.E., 2005. Containing Pandemic Influenza at the Source. *Science*, 309(5737), pp.1083-1087.

Maier, B.F., and Brockmann, D., 2020. Effective Containment Explains Subexponential Growth in Recent Confirmed COVID-19 Cases in China. *Science*, 368(6492), pp.742-746.

Manning, A. and Petrongolo, B., 2017. How Local are Labor Markets? Evidence from a Spatial Job Search Model. *American Economic Review*, 107(10), pp.2877-2907.

Miller K, Curry K. 2020. The COVID Tracking Project. https://github.com/COVID19Tracking

Monte, F., Redding, S.J., and Rossi-Hansberg, E., 2018. Commuting, Migration, and Local Employment Elasticities. *American Economic Review*, 108(12), pp.3855-90.

Piguillem, F., and Shi, L., 2022. Optimal COVID-19 Quarantine and Testing Policies. *The Economic Journal*, 132(64), pp.2534–2562.

Redding, S.J., and Rossi-Hansberg, E., 2017. Quantitative Spatial Economics. *Annual Review of Economics*, 9, pp.21-58.

Rogerson, R., Shimer, R. and Wright, R., 2005. Search-theoretic Models of the Labor Market: A Survey. *Journal of Economic Literature*, 43(4), pp.959-988.

Schmutz, B. and Sidibé, M., 2019. Frictional Labour Mobility. *The Review of Economic Studies*, 86(4), pp.1779-1826.

Silva, J.S., and Tenreyro, S., 2006. The Log of Gravity. *The Review of Economics and Statistics*, 88(4), pp.641-658.

Tombe, T., and Zhu, X., 2019. Trade, Migration, and Productivity: A Quantitative Analysis of China. *American Economic Review*, 109(5), pp.1843-72.

Van den Driessche, P., 2017. Reproduction Numbers of Infectious Disease Models. *Infectious Disease Modelling*, 2(3), pp.288-303.

Waugh, M.E., 2010. International Trade and Income Differences. *American Economic Review*, 100(5), pp.2093-2124.

Endogenous Mobility in Pandemics: Theory and Evidence from the United States

(For Online Publication)

by

Xiao Chen Hanwei Huang Jiandong Ju
Ruoyan Sun Jialiang Zhang

Appendix A: Proofs

A.1.1 Proof of Proposition 1

We start from the value function and mobility matrix of the recovered individuals. Taking expectation on both sides of equation (7) and denote $\widetilde{V}_{i,t}^R \equiv E_{t-1}U_{i,t}^R(\boldsymbol{\varepsilon_{i,t}})$, we have

$$\widetilde{V}_{i,t}^R = u_i + E_{\boldsymbol{\varepsilon_{i,t}}} \max_j \{\beta \widetilde{V}_{j,t+1}^R - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\}. \tag{E1}$$

Let's denote $V \equiv \max_j \{\beta \widetilde{V}_{j,t+1}^R - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\}$, we can calculate its expectation under our assumption that $\varepsilon_{ij,t}$ follows the Gumbel distribution:

$$\begin{aligned}
E_{\boldsymbol{\varepsilon_{i,t}}} V &= \int V \mathrm{d} Pr\{\forall j, \varepsilon_{ij,t} \leq V + \widetilde{\mu}_{ij,t} - \beta \widetilde{V}_{j,t+1}^R\} \\
&= \int V \mathrm{d} \prod_j \exp\left\{-\exp\left\{-[V + \widetilde{\mu}_{ij,t} - \beta \widetilde{V}_{j,t+1}^R]/\kappa - \gamma^{Euler}\right\}\right\} \\
&= \int V \mathrm{d} \exp\left\{-\sum_j (V_{j,t+1}^R)^\beta \mu_{ij,t}^{-1} \exp\left\{-V/\kappa - \gamma^{Euler}\right\}\right\}.
\end{aligned}$$

Let $Z \equiv \kappa \ln \sum_j (V_{j,t+1}^R)^\beta \mu_{ij,t}^{-1}$, then the integral above can be transformed as:

$$\begin{aligned}
E_{\boldsymbol{\varepsilon_{i,t}}} V &= \int V \mathrm{d} \exp\left\{-\exp\left\{-(V - Z)/\kappa - \gamma^{Euler}\right\}\right\} \\
&= \int (V' + Z) \mathrm{d} \exp\left\{-\exp\left\{-V'/\kappa - \gamma^{Euler}\right\}\right\} = Z,
\end{aligned}$$

where the last equality is due to our normalization of the mean of the Gumbel distribution as 0. Substituting the result above to equation (E1), we get

$$\widetilde{V}_{i,t}^R = u_i + E_{\boldsymbol{\varepsilon_{i,t}}} V \Rightarrow V_{i,t}^R = \exp\left(\frac{u_i}{\kappa}\right) \sum_{j=1}^N \left(V_{j,t+1}^R\right)^\beta \left(\mu_{ij,t}\right)^{-1}.$$

The value functions of type S and type I can be derived in similar ways.

A.1.2 Proof of Proposition 2

To simply notations, we denote $Z_{ij} = \beta\widetilde{V}^R_{j,t+1} - \widetilde{\mu}_{ij,t}$, the mobility rate from i to j for type R satisfies:

$$\begin{aligned} m^R_{ij,t} &= Prob\{\forall k \neq j, \beta\widetilde{V}^R_{k,t+1} - \widetilde{\mu}_{ik,t} + \varepsilon_{ik,t} \leq \beta\widetilde{V}^R_{j,t+1} - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\} \\ &= Prob\{\forall k \neq j, \varepsilon_{ik,t} \leq \varepsilon_{ij,t} + Z_{ij} - Z_{ik}\} \\ &= \int \prod_{k\neq j} \exp\left\{-\exp\left\{-(\varepsilon_{ij,t} + Z_{ij} - Z_{ik})/\kappa - \gamma^{Euler}\right\}\right\} \mathrm{d}\exp\left\{-\exp\left\{-\varepsilon_{ij,t}/\kappa - \gamma^{Euler}\right\}\right\}. \end{aligned}$$

Denote $Y = \exp\left\{-\exp\left\{-\varepsilon_{ij,t}/\kappa - \gamma^{Euler}\right\}\right\}$, we have:

$$\begin{aligned} m^R_{ij,t} &= \int Y^{\sum_{k\neq j}\exp\{(Z_{ik}-Z_{ij})/\kappa\}}\mathrm{d}Y = \frac{1}{\sum_{k\neq j}\exp\{(Z_{ik}-Z_{ij})/\kappa\}+1} \\ &= \frac{\exp\{Z_{ij}/\kappa\}}{\sum_k \exp\{Z_{ik}/\kappa\}} = \frac{\left(V^R_{j,t+1}\right)^\beta (\mu_{ij,t})^{-1}}{\sum_{k=1}^N \left(V^R_{k,t+1}\right)^\beta (\mu_{ik,t})^{-1}}. \end{aligned}$$

The mobility rates for type S and I can be derived in similar ways.

A.1.3 Proof of Proposition 3

Denote $\widehat{x} = x/x_{ss}$, while x_{ss} is the steady-state value of x, we have $\hat{V}^R_{i,t}$ and $\hat{V}^I_{i,t}$ converging to 1 at the steady state. It implies that $\forall\delta > 0$ small enough, $\exists\, T$ such that $\forall t > T$, $|\hat{V}^R_{i,t} - 1| < \delta$. Using the Taylor expansion, we know that $|\left(\hat{V}^R_{i,T+1}\right)^\beta - 1| < \beta\delta + o(\delta)$.

Suppose for any period s, $\exists\,\epsilon$ such that $|\hat{V}^R_{i,s} - 1| = \epsilon > 0$, from equations (8) and (9), we know

$$\begin{aligned} \hat{V}^R_{i,T} - 1 &= \sum_{j=1}^N m^R_{ij,ss}\left[\left(\hat{V}^R_{j,T+1}\right)^\beta - 1\right] \\ \Leftrightarrow |\hat{V}^R_{i,T} - 1| &< \sum_{j=1}^N m^R_{ij,ss}(\beta\delta + o(\delta)) = \beta\delta + o(\delta) \\ \Leftrightarrow |\hat{V}^R_{i,s} - 1| &< \beta^{T+1-s}\delta + o(\delta) \end{aligned}$$

Then for the given δ, since $\beta < 1$ we can always find some T large enough such that

$$|\hat{V}^R_{i,s} - 1| < \beta^{T+1-s}\delta + o(\delta) < \epsilon, \tag{E2}$$

contradiction, we have $\hat{V}_{i,t}^R = 1, \forall t$. Similarly, we can show that $\hat{V}_{i,t}^I = 1$. When the value functions of type R and I, and mobility costs μ_t are time-invariant, it is easy to see that $m_{ij,t}^R = m_{ij}^R$ and $m_{ij,t}^I = m_{ij}^I$ from equation (9).

A.1.4 Proof of Lemma 1

Proposition 3 tells that the lifetime utility types R and I are time-invariant. And if mobility cost μ_t is also time-invariant, equation (8) becomes

$$
\begin{aligned}
V_{i,t}^S &= e^{u_i/\kappa} \sum_{j=1}^N (V_{j,t+1}^S)^{\beta(1-\alpha_{j,t+1})} (V_j^I)^{\beta\alpha_{j,t+1}} \mu_{ij}^{-1} \\
V_i^I &= e^{u_i/\kappa} \sum_{j=1}^N (V_j^I)^{\beta(1-\gamma_j^R-\gamma_j^D)} (V_j^R)^{\beta\gamma_j^R} (V_j^D)^{\beta\gamma_j^D} \mu_{ij}^{-1} \\
V_i^R &= e^{u_i/\kappa} \sum_{j=1}^N (V_j^R)^{\beta} \mu_{ij}^{-1}
\end{aligned}
$$

Divide the equation of type I by the one of type R, and use equation (9), we get:

$$
\frac{V_i^I}{V_i^R} = \sum_{j=1}^N m_{ij}^R \left(\frac{V_j^I}{V_j^R}\right)^{\beta(1-\gamma_j^R-\gamma_j^D)} \left(\frac{V_j^D}{V_j^R}\right)^{\beta\gamma_j^D} < \sum_{j=1}^N m_{ij}^R \left(\frac{V_j^I}{V_j^R}\right)^{\beta(1-\gamma_j^R-\gamma_j^D)},
$$

where we use the assumption $V_i^D < V_i^R, \forall i$. Denote $\overline{\lambda} \equiv \max_i\{(\frac{V_i^I}{V_i^R})^{\beta(1-\gamma_i^R-\gamma_i^D)}\}$ and $k \equiv \arg\max_i\{(\frac{V_i^I}{V_i^R})^{\beta(1-\gamma_i^R-\gamma_i^D)}\}$ and note that $\beta(1-\gamma_k^D-\gamma_k^R) < 1$, the inequality above implies that

$$
\overline{\lambda}^{\frac{1}{\beta(1-\gamma_k^R-\gamma_k^D)}} < \overline{\lambda} \Rightarrow \overline{\lambda} < 1 \Rightarrow V_i^I < V_i^R, \quad \forall i.
$$

Similarly, we divide the equation of type S group by the one of type R and get:

$$
\frac{V_{i,t}^S}{V_i^R} = \sum_{j=1}^N m_{ij}^R \left(\frac{V_{j,t+1}^S}{V_j^R}\right)^{\beta(1-\alpha_{j,t+1})} \left(\frac{V_j^I}{V_j^R}\right)^{\beta\alpha_{j,t+1}} \leq \sum_{j=1}^N m_{ij}^R \left(\frac{V_{j,t+1}^S}{V_j^R}\right)^{\beta(1-\alpha_{j,t+1})}
$$

Say, if there exists some $V_{i,t}^S/V_i^R > 1$, from the inequality above, we know that there must exist some $V_{j,t+1}^S/V_j^R > 1$. Suppose $\overline{\xi}_t \equiv \max_i V_{i,t}^S/V_i^R > 1$, then the equation above implies that

$$
\overline{\xi}_t \leq \sum_{j=1}^N m_{ij}^R \overline{\xi}_{t+1}^{\beta(1-\alpha_{j,t+1})} \leq \sum_{j=1}^N m_{ij}^R \overline{\xi}_{t+1}^{\beta} = \overline{\xi}_{t+1}^{\beta}
$$

Therefore, $\bar{\xi}_{t+1} \geq \bar{\xi}_t^{1/\beta} > \bar{\xi}_t$ and $\{\bar{\xi}_{t+1}\}_{t=0}^{\infty}$ is a growing series greater than 1. However, at the steady state, as the infected individuals are all removed from the population and infection rates fall to zero, we have $V_{i,\infty}^S = V_{i,\infty}^R$ according to equation (8). We should have $\bar{\xi}_\infty = 1$. Contradiction, we have $V_{i,t}^S \leq V_{i,t}^R, \ \forall i$ and t. Finally, near DFE, the measure of type I is close to 0 and the whole population is susceptible, we have $V_{i,t}^S = V_i^R$ according to equation (8). Since $V_i^I < V_i^R$, we immediately have $V_i^I < V_{i,t}^S$ near the DFE.

A.1.5 Proof of Proposition 4

Taking logs on both sides of the expression for $m_{ij,t}^S$ in equation (9), we have

$$\ln(m_{ij,t}^S) = \beta(1-\alpha_{j,t+1})\ln(V_{j,t+1}^S) + \beta\alpha_{j,t+1}\ln(V_{j,t+1}^I) - \ln(\mu_{ij,t}) - ln(OV_{i,t}^S), \tag{E3}$$

while $OV_{i,t}^S \equiv \sum_{k=1}^N \left(V_{k,t+1}^S\right)^{\beta(1-\alpha_{k,t+1})} \left(V_{k,t+1}^I\right)^{\beta\alpha_{k,t+1}} (\mu_{ik,t})^{-1}$ is the option value of region i for type S. Taking partial derivative of equation (E3) with respect to $\alpha_{j,t+1}$, we have

$$\frac{\partial \ln(m_{ij,t}^S)}{\partial \alpha_{j,t+1}} = \beta \ln(\frac{V_{j,t+1}^I}{V_{j,t+1}^S}) - \frac{\partial \ln(OV_{i,t}^S)}{\partial \alpha_{j,t+1}}. \tag{E4}$$

Denote $v_{k,t} = \left(V_{k,t+1}^S\right)^{\beta(1-\alpha_{k,t+1})} \left(V_{k,t+1}^I\right)^{\beta\alpha_{k,t+1}} (\mu_{ik,t})^{-1}$, then $OV_{i,t}^S = \sum_{k=1}^N v_{k,t}$ and

$$\begin{aligned}\frac{\partial \ln(OV_{i,t}^S)}{\partial \alpha_{j,t+1}} &= \frac{\partial \ln(OV_{i,t}^S)}{\partial OV_{i,t}^S}\frac{\partial OV_{i,t}^S}{\partial v_{j,t}}\frac{\partial v_{j,t}}{\partial ln(v_{j,t})}\frac{\partial \ln(v_{j,t})}{\partial \alpha_{j,t+1}} = \frac{v_{j,t}}{OV_{i,t}^S}\beta\ln(\frac{V_{j,t+1}^I}{V_{j,t+1}^S}) \\ &= \beta m_{ij,t}^S \ln(\frac{V_{j,t+1}^I}{V_{j,t+1}^S}).\end{aligned}$$

The last equality uses the expression for $m_{ij,t}^S$ in equation (9). Therefore, we have

$$\frac{\partial \ln(m_{ij,t}^S)}{\partial \alpha_{j,t+1}} = \beta(1-m_{ij,t}^S)\ln(\frac{V_{j,t+1}^I}{V_{j,t+1}^S}), \tag{E5}$$

Since $V_j^I < V_{j,t}^S$ according to Lemma 1 and $0 < m_{ij,t}^S < 1$ we have $\frac{\partial \ln(m_{ij,t}^S)}{\partial \alpha_{j,t+1}} < 0$.

A.1.6 Proof of Proposition 5

We know that the law of motion for type I is given by $\boldsymbol{I_{t+1}} = \boldsymbol{F}\boldsymbol{I_t} + \boldsymbol{V}\boldsymbol{I_t}$, and the next-generation matrix $\boldsymbol{m}^I$ is $\boldsymbol{m}^I = \boldsymbol{F}(\mathbb{I} - \boldsymbol{V})^{-1}$. Given that $\chi_i \geq 0$ and $m_{ij}^I \geq 0$, $\forall i$ and $\forall j$, $\mathbf{F}$ is obviously non-negative. As $\mathbf{V}$ has a spectral radius of matrix less than 1, we have $(\mathbb{I} - \boldsymbol{V})^{-1} = \sum_{t=1}^{\infty} \mathbf{V}^t$. Since $\mathbf{V}$ is non-negative, as $\gamma_i < 1 \ \forall i$, $(\mathbb{I} - \boldsymbol{V})^{-1}$ must be non-negative as well. Therefore, $\mathbf{M}$ is

non-negative. Then using the *Collatz - Wielandt* formula, the spectral radius of $\boldsymbol{m}^I$ satisfies:

$$\rho(\boldsymbol{m}^I) = \max_{\boldsymbol{x}\geq 0} \min_{i,x_i\neq 0} \frac{(\boldsymbol{xM})_i}{x_i} = \min_{\boldsymbol{x}\geq 0} \max_{i,x_i\neq 0} \frac{(\boldsymbol{xM})_i}{x_i}. \tag{E6}$$

Denote $\boldsymbol{e} = [1, ..., 1]_{1\times N}$, a row vector that all elements are one. According to the *Perron-Frobenius* theorem, $\boldsymbol{e}$ is the left eigenvector of the mobility matrix $\boldsymbol{m}^I$ with an eigenvalue of 1, given that $\boldsymbol{m}^I$ is a Markov matrix. So we have $\boldsymbol{em^I} = \boldsymbol{e}$. We can construct a row vector $\boldsymbol{\varepsilon}$ whose i^{th} element is $\varepsilon_i = \gamma_i/\chi_i$, i.e., $\boldsymbol{\varepsilon} = [\gamma_1/\chi_1, ..., \gamma_N/\chi_N]_{1\times N}$. We next show

$$\boldsymbol{\varepsilon M} = \boldsymbol{\varepsilon F}(\mathbb{I} - \boldsymbol{V})^{-1} = \boldsymbol{e}.$$

First, let $Diag(\boldsymbol{x})$ be a square matrix with elements of vector $\boldsymbol{x}$ on the main diagonal, and $\boldsymbol{\chi} = \{\chi_1, ..., \chi_N\}$ and $\boldsymbol{\gamma} = \{\gamma_1, ..., \gamma_N\}$. Then $\boldsymbol{\varepsilon F} = \boldsymbol{\varepsilon} Diag(\boldsymbol{\chi})\boldsymbol{m}^I = \boldsymbol{\gamma m}^I$. Second, $\boldsymbol{e}(\mathbb{I} - \boldsymbol{V}) = \boldsymbol{e}(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I) = \boldsymbol{e} - \boldsymbol{e}Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I = \boldsymbol{em^I} - (\boldsymbol{em^I} - \boldsymbol{e}Diag(\boldsymbol{\gamma})\boldsymbol{m}^I) = \boldsymbol{\gamma m}^I$. Therefore, $\boldsymbol{\varepsilon F} = \boldsymbol{e}(\mathbb{I} - \boldsymbol{V})$, or equivalently, $\boldsymbol{\varepsilon F}(\mathbb{I} - \boldsymbol{V})^{-1} = \boldsymbol{e}$. Then if we replace $\boldsymbol{x}$ in equation (E6) by the constructed vector $\boldsymbol{\varepsilon}$, we have

$$\min_i \frac{\chi_i}{\gamma_i} = \min_i \frac{1}{\varepsilon_i} \leq \rho(\boldsymbol{m}^I) \leq \max_i \frac{1}{\varepsilon_i} = \max_i \frac{\chi_i}{\gamma_i}, \tag{E7}$$

so the global R_0 is bounded by the lowest and highest local R_0.

A.1.7 Proof of Proposition 6

Following the previous proof, we denote the next-generation matrix as $\boldsymbol{M} = \boldsymbol{F}(\mathbb{I} - \boldsymbol{V})^{-1} = Diag(\boldsymbol{\chi})\boldsymbol{m}^I(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1}$, therefore,

$$\frac{\mathrm{d}\boldsymbol{m}^I}{\mathrm{d}\chi_i} = \frac{\mathrm{d}Diag(\boldsymbol{\chi})}{\mathrm{d}\chi_i}\boldsymbol{m}^I(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1} > \boldsymbol{0},$$

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{m}^I}{\mathrm{d}\gamma_i} &= -Diag(\boldsymbol{\chi})\boldsymbol{m}^I(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1}\frac{(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)}{\mathrm{d}\gamma_i}(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1} \\ &= Diag(\boldsymbol{\chi})\boldsymbol{m}^I(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1}\frac{\mathrm{d}Diag(\boldsymbol{e} - \boldsymbol{\gamma})}{\mathrm{d}\gamma_i}\boldsymbol{m}^I(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1} < \boldsymbol{0},\end{aligned}$$

using the fact that matrices $\boldsymbol{m}^I$ and $(\mathbb{I} - Diag(\boldsymbol{e} - \boldsymbol{\gamma})\boldsymbol{m}^I)^{-1}$ are positive matrices. According to Theorem 2.1 (c) in Li and Schneider (2002), the spectral radius of an irreducible nonnegative matrix increases if any entry increases, and decreases if any entry decreases. Hence, the global R_0 increases with the transmission rate χ_i and decreases with the removal rate $\gamma_i, \forall\ i$.

A.1.8 Proof of Proposition 7

We divide the value function for type S in equation (8) by its steady state and have

$$\begin{aligned}\hat{V}_{i,t}^{S} = \frac{V_{i,t}^{S}}{V_{i,ss}^{S}} &= \frac{\sum_{j=1}^{N}\left(V_{j,t+1}^{S}\right)^{\beta(1-\alpha_{j,t+1})}\left(V_{j,t+1}^{I}\right)^{\beta\alpha_{j,t+1}}\left(\mu_{ij,t}\right)^{-1}}{\sum_{j=1}^{N}\left(V_{j,ss}^{S}\right)^{\beta}\left(\mu_{ij,ss}\right)^{-1}} \\ &= \sum_{j=1}^{N}\frac{\left(V_{j,ss}^{S}\right)^{\beta}\left(\mu_{ij,ss}\right)^{-1}}{\sum_{k=1}^{N}\left(V_{k,ss}^{S}\right)^{\beta}\left(\mu_{ik,ss}\right)^{-1}}\frac{\left(V_{j,t+1}^{S}\right)^{\beta(1-\alpha_{j,t+1})}\left(V_{j,t+1}^{I}\right)^{\beta\alpha_{j,t+1}}\left(\mu_{ij,t}\right)^{-1}}{\left(V_{j,ss}^{S}\right)^{\beta}\left(\mu_{ij,ss}\right)^{-1}} \\ &= \sum_{j=1}^{N} m_{ij,ss}^{S}\left(\hat{V}_{j,t+1}^{S}\right)^{\beta(1-\alpha_{j,t+1})}\left(\hat{V}_{j,t+1}^{I}\right)^{\beta\alpha_{j,t+1}}\left(\frac{V_{j,ss}^{I}}{V_{j,ss}^{S}}\right)^{\beta\alpha_{j,t+1}}\left(\widehat{\mu}_{ij,t}\right)^{-1}.\end{aligned}$$

Similarly,

$$\begin{aligned}\hat{V}_{i,t}^{I} &= \frac{V_{i,t}^{I}}{V_{i,ss}^{I}} = \frac{\exp\left(\frac{u_i}{\kappa}\right)\sum_{j=1}^{N}\left(V_{j,t+1}^{I}\right)^{\beta(1-\gamma_j^R-\gamma_j^D)}\left(V_{j,t+1}^{R}\right)^{\beta\gamma_j^R}\left(V^D\right)^{\gamma_j^D}\left(\mu_{ij,t}\right)^{-1}}{\exp\left(\frac{u_i}{\kappa}\right)\sum_{j=1}^{N}\left(V_{j,ss}^{I}\right)^{\beta(1-\gamma_j^R-\gamma_j^D)}\left(V_{j,ss}^{R}\right)^{\beta\gamma_j^R}\left(V^D\right)^{\gamma_j^D}\left(\mu_{ij,ss}\right)^{-1}} \\ &= \frac{\sum_{k=1}^{N}\left(V_{k,t+1}^{I}\right)^{\beta(1-\gamma_k^R-\gamma_k^D)}\left(V_{k,t+1}^{R}\right)^{\beta\gamma_k^R}\left(V^D\right)^{\gamma_k^D}\left(\mu_{ik,t}\right)^{-1}}{\sum_{j=1}^{N}\left(V_{j,ss}^{I}\right)^{\beta(1-\gamma_j^R-\gamma_j^D)}\left(V_{j,ss}^{R}\right)^{\beta\gamma_j^R}\left(V^D\right)^{\gamma_j^D}\left(\mu_{ij,ss}\right)^{-1}} \\ &= \sum_{k=1}^{N}\frac{\left(V_{k,ss}^{I}\right)^{\beta(1-\gamma_k^R-\gamma_k^D)}\left(V_{k,ss}^{R}\right)^{\beta\gamma_k^R}\left(V^D\right)^{\gamma_k^D}\left(\mu_{ik,ss}\right)^{-1}}{\sum_{j=1}^{N}\left(V_{j,ss}^{I}\right)^{\beta(1-\gamma_j^R-\gamma_j^D)}\left(V_{j,ss}^{R}\right)^{\beta\gamma_j^R}\left(V^D\right)^{\gamma_j^D}\left(\mu_{ij,ss}\right)^{-1}}\left(\widehat{V}_{k,t+1}^{I}\right)^{\beta(1-\gamma_k^R-\gamma_k^D)}\left(\widehat{V}_{k,t+1}^{R}\right)^{\beta\gamma_k^R}\widehat{\mu}_{ik,t}^{-1} \\ &= \sum_{k=1}^{N} m_{ik,ss}^{I}\left(\widehat{V}_{k,t+1}^{I}\right)^{\beta(1-\gamma_k^R-\gamma_k^D)}\left(\widehat{V}_{k,t+1}^{R}\right)^{\beta\gamma_k^R}\widehat{\mu}_{ik,t}^{-1},\end{aligned}$$

$$\begin{aligned}\hat{V}_{i,t}^{R} &= \frac{V_{i,t}^{R}}{V_{i,ss}^{R}} = \frac{\exp\left(\frac{u_i}{\kappa}\right)\sum_{j=1}^{N}\left(V_{j,t+1}^{R}\right)^{\beta}\left(\mu_{ij,t}\right)^{-1}}{\exp\left(\frac{u_i}{\kappa}\right)\sum_{j=1}^{N}\left(V_{j,ss}^{R}\right)^{\beta}\left(\mu_{ij,ss}\right)^{-1}} = \frac{\sum_{k=1}^{N}\left(V_{k,t+1}^{R}\right)^{\beta}\left(\mu_{ik,t}\right)^{-1}}{\sum_{j=1}^{N}\left(V_{j,ss}^{R}\right)^{\beta}\left(\mu_{ij,ss}\right)^{-1}} \\ &= \sum_{k=1}^{N}\frac{\left(V_{k,ss}^{R}\right)^{\beta}\left(\mu_{ik,ss}\right)^{-1}}{\sum_{j=1}^{N}\left(V_{j,ss}^{R}\right)^{\beta}\left(\mu_{ij,ss}\right)^{-1}}\frac{\left(V_{k,t+1}^{R}\right)^{\beta}\left(\mu_{ik,t}\right)^{-1}}{\left(V_{k,ss}^{R}\right)^{\beta}\left(\mu_{ik,ss}\right)^{-1}} \\ &= \sum_{k=1}^{N} m_{ik,ss}^{R}\left(\hat{V}_{i,t+1}^{R}\right)^{\beta}\widehat{\mu}_{ik,t}^{-1}.\end{aligned}$$

As for the mobility rate, we do the same and get

$$
\begin{aligned}
\frac{m^S_{ij,t}}{m^S_{ij,ss}} &= \frac{\left(V^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})}\left(V^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}}(\mu_{ij,t})^{-1}}{\sum_{k=1}^N\left(V^S_{k,t+1}\right)^{\beta(1-\alpha_{k,t+1})}\left(V^I_{k,t+1}\right)^{\beta\alpha_{k,t+1}}(\mu_{ik,t})^{-1}}\frac{\sum_{k=1}^N\left(V^S_{k,ss}\right)^{\beta}(\mu_{ik,ss})^{-1}}{\left(V^S_{j,ss}\right)^{\beta}(\mu_{ij,ss})^{-1}} \\
&= \frac{\left(V^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})}\left(V^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}}(\mu_{ij,t})^{-1}}{\left(V^S_{j,ss}\right)^{\beta}(\mu_{ij,ss})^{-1}}\frac{V^S_{i,ss}}{V^S_{i,t+1}} \\
&= \frac{\left(V^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})}\left(V^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}}(\mu_{ij,t})^{-1}}{\left(V^S_{j,ss}\right)^{\beta}(\mu_{ij,ss})^{-1}}\frac{1}{\hat{V}^S_{i,t+1}} \\
&= \frac{\left(\hat{V}^S_{j,t+1}\right)^{\beta(1-\alpha_{j,t+1})}\left(\hat{V}^I_{j,t+1}\right)^{\beta\alpha_{j,t+1}}\left(\frac{V^I_{j,ss}}{V^S_{j,ss}}\right)^{\beta\alpha_{j,t+1}}(\widehat{\mu}_{ij,t})^{-1}}{\sum_{k=1}^N m^S_{ik,ss}\left(\hat{V}^S_{k,t+1}\right)^{\beta(1-\alpha_{k,t+1})}\left(\hat{V}^I_{k,t+1}\right)^{\beta\alpha_{k,t+1}}\left(\frac{V^I_{k,ss}}{V^S_{k,ss}}\right)^{\beta\alpha_{k,t+1}}(\widehat{\mu}_{ik,t})^{-1}}.
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
\widehat{m}^I_{ij,t} &= \frac{\left(V^I_{j,t+1}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)}\left(V^R_{j,t+1}\right)^{\beta\gamma^R_j}\left(V^D\right)^{\beta\gamma^D_j}(\mu_{ij,t})^{-1}}{\left(V^I_{j,ss}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)}\left(V^R_{j,ss}\right)^{\beta\gamma^R_j}\left(V^D\right)^{\beta\gamma^D_j}(\mu_{ij,ss})^{-1}}\frac{V^I_{i,ss}}{V^I_{i,t}} \\
&= \frac{1}{\hat{V}^I_{i,t}}\left(\hat{V}^I_{i,t+1}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)}\left(\hat{V}^R_{j,t+1}\right)^{\beta\gamma^R_j}(\widehat{\mu}_{ij,t})^{-1} \\
&= \frac{\left(\hat{V}^I_{i,t+1}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)}\left(\hat{V}^R_{j,t+1}\right)^{\beta\gamma^R_j}(\widehat{\mu}_{ij,t})^{-1}}{\sum_{k=1}^N m^I_{ik,ss}\left(\widehat{V}^I_{k,t+1}\right)^{\beta(1-\gamma^R_k-\gamma^D_k)}\left(\widehat{V}^R_{k,t+1}\right)^{\beta\gamma^R_k}\widehat{\mu}^{-1}_{ik,t}},
\end{aligned}
$$

$$
\begin{aligned}
\widehat{m}^R_{ij,t} &= \frac{\left(V^R_{j,t+1}\right)^{\beta}(\mu_{ij,t})^{-1}}{\left(V^R_{j,ss}\right)^{\beta}(\mu_{ij,ss})^{-1}}\frac{V^R_{i,ss}}{V^R_{i,t}} = \left(\widehat{V}^R_{j,t+1}\right)^{\beta}\widehat{\mu}^{-1}_{ij,t}\frac{1}{\widehat{V}^R_{i,t}} \\
&= \frac{\left(\widehat{V}^R_{j,t+1}\right)^{\beta}\widehat{\mu}^{-1}_{ij,t}}{\sum_{k=1}^N m^R_{ik,ss}\left(\hat{V}^R_{i,t+1}\right)^{\beta}\widehat{\mu}^{-1}_{ik,t}}.
\end{aligned}
$$

Appendix B: Supplementary Materials

B.1 Constructing the Mobility Matrix using PlaceIQ Data

We use the first bi-week *PlaceIQ* bilateral mobility probability matrix from January 2020 to conduct estimation, where the row indicates the origin and the column indicates the destination. However, there are two problems regarding the raw data. First, the sum of mobility probabilities from all origins towards a specific destination generally exceeds one, which may be the consequence of multiple counts for a resident crossing borders several times within a period. Second, the mobility probabilities calculated in raw data are denominated by device counts in destination instead of origin, where only the latter method accords with our model definition.

We construct the mobility matrix using *PlaceIQ* data. In the first step, we rearrange the mobility matrix to guarantee unit column sum by keeping the diagonal entries unchanged while scaling the off-diagonal entries proportionally. If we neglect the cases where residents cross borders at least three times in two weeks, multiple counting only occurs when a resident in region i moves out to region j, and then back to i or to another region k. So the probability that region i resident stays in region i is not biased, but the frequencies that region i resident moves to other regions is over-presented. As long as the bilateral over-estimated cases are proportional to the total observed bilateral mobility raw data, scaling the off-diagonal column elements generates the unit sum matrix with little bias. In the second step, we transform the mobility probability matrix by multiplying each column by the destination population and then scaling it by the row rums to keep it consistent with the model definition.

B.2 Additional Tables

Table B.1: COVID-19 Containment Policies

ID	Description	Coding
C1	Closings of schools and universities	0 - no measures 1 - recommend closing or all schools open with alterations 2 - require closing (only some levels or categories, eg just high school) 3 - require closing all levels
C2	Closings of workplaces	0 - no measures 1 - recommend closing or all businesses open with alterations 2 - require closing (or work from home) for some sectors or categories of workers 3 - require closing (or work from home) for all-but-essential workplaces
C3	Cancelling of public events	0 - no measures 1 - recommend canceling 2 - require cancelling
C4	Cut-off size for limits on gatherings	0 - no restrictions 1 - restrictions on very large gatherings (the limit is above 1000 people) 2 - restrictions on gatherings between 101-1000 people 3 - restrictions on gatherings between 11-100 people 4 - restrictions on gatherings of 10 people or less
C5	Closing of public transport	0 - no measures 1 - recommend closing (or significantly reduce volume/route/means of transport) 2 - require closing (or prohibit most citizens from using it)
C6	Orders to ”shelter-in-place” and otherwise confine to the home	0 - no measures 1 - recommend not leaving house 2 - require not leaving house with exceptions for ’essential’ trips 3 - require not leaving house with minimal exceptions
C7	Restrictions on internal movement between cities/regions	0 - no measures 1 - recommend not to travel between regions/cities 2 - internal movement restrictions in place
C8	Restrictions on international travel	0 - no restrictions 1 - screening arrivals 2 - quarantine arrivals from some or all regions 3 - ban arrivals from some regions 4 - ban on all regions or total border closure

Notes. This table lists policies that are classified as containment policies in the Oxford COVID-19 Government Response Tracker (Hale et al., 2021).

Table B.2: Summary Statistics

Variables	Definition	Obs	Mean	Std.Dev	Min	Max
Panel A. Variables at the bilateral state pair - biweek level						
max mobility	the maximum of bilateral mobility rate between states	102,000	0.012	0.036	0	0.766
mean mobility	average bilateral mobility rate between states	102,000	0.011	0.034	0	0.742
distance	bilateral distance between states (in km)	102,000	1,981.0	1,461.3	37.4	8,176.1
border	a dummy variable indicating whether two states share a border	102,000	0.084	0.277	0	1
Panel B. Variables at the state - bi-week level						
new infection rate	new cases/population at state-bi-week level	2,040	0.003	0.003	0	0.024
mobility restrictions	restrictions on internal travel across regions/cities implemented by state government	2,040	1.116	0.537	0	2
stringency index	US state-level containment policy index, ranging from 0 to 100	2,040	52.4	18.300	0	93.5
number of new cases	number of new cases at state bi-weekly level	2,040	17,930	37,035	0	575,710
$\sum_{j\neq i} m_{ji}^{max} ln(newcases_{j,t})$	average of new cases (in logs) across other states weighted by max initial mobility	2,040	13.3	5.713	0	26.96
$\sum_{j\neq i} m_{ji}^{mean} ln(newcases_{j,t})$	average of new cases (in logs) across other states weighted by mean initial mobility	2,040	12.91	5.50	0	26.04
$\sum_{j\neq i} m_{ji}^{max} ln(cumulativecases_{j,t})$	average of cumulative cases (in logs) across other states weighted by max initial mobility	2,040	17.39	7.49	0.001	33.0
$\sum_{j\neq i} m_{ji}^{mean} ln(cumulativecases_{j,t})$	average of cumulative cases (in logs) across other states weighted by mean initial mobility	2,040	16.8	7.2	0.001	31.7
max net inflow	total net inflow (maximum within a bi-week) over the local population	2,040	0.136	3.43	-10.2	11.80
mean net inflow	total net inflow (average within a bi-week) over the local population	2,040	0.121	3.156	-9.67	11.02
cumulative infection rate	cumulative number of cases over the local population at the end of the sample period	2,040	10.98	2.58	3.55	14.91

Notes. Panel A shows the summary statistics of variables at the bilateral state pair-biweek level. The sample includes all 2,550 state pairs in the US, from January 2020 to August 2021, with 40 periods of bi-weeks. Panel B shows the summary statistics of variables at the state-biweek level. The sample includes all 51 states in the US, from January 2020 to August 2021, with 40 periods of bi-weeks.

Table B.3: Model Parameters for the Three-Region Economy

Parameters	Definition	Value
β	Discount rate	0.8
κ	Dispersion of preference shock	8
γ^D	Mortality rate	0.3
γ^R	Recovery rate	0.4
χ_i	Transmission rate of Region i	1.2
μ_{ij}	mobility costs from Region i to Region j, $i \neq j$	exp(2)
μ_{ii}	local mobility costs	1
u_i	instantaneous utility of types S, I and R	0
u^D	instantaneous utility of type D	$\ln(e^{-15})$

Notes. This table shows the parameters specified for the three-region economy simulated in Section 4.2.

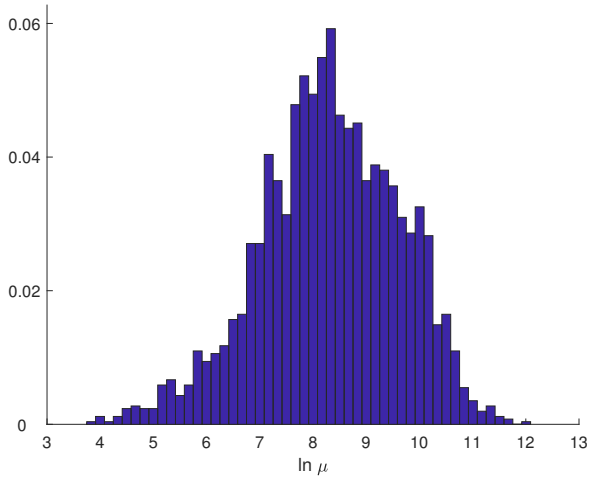
Table B.4: Robustness Checks on Counterfactual Simulations

	Alternative parameters	Scenarios	Total Infection (millions)	Total deaths (millions)	Welfare relative to *Baseline*
(**a**)	$\beta = 0.96$	*Baseline*	89.6	2.63	
		Exogenous mobility	113.4	5.78	-20.70%
		No mobility control	98.8	2.31	4.67%
		No local control	144.5	3.97	-49.16%
(**b**)	$\ln V^D = -10$	*Baseline*	82.4	3.33	
		Exogenous mobility	113.4	5.78	-18.35%
		No mobility control	87.4	1.87	70.41%
		No local control	144.6	4.75	-2.95%
(**c**)	$\gamma^D = 0.0164$	*Baseline*	81.6	4.22	
		Exogenous mobility	112.9	5.57	-18.44%
		No mobility control	83.0	1.88	71.89%
		No local control	144.9	5.57	-3.02%
(**d**)	$\gamma^R = 0.697$	*Baseline*	89.4	1.90	
		Exogenous mobility	104.2	2.23	-18.17%
		No mobility control	102.4	2.16	65.62%
		No local control	148.4	3.27	-2.73%

Notes: This table conducts robustness checks using alternative parameters, changing one parameter at a time. Row (a) replaces the discount factor β from 0.9985 to 0.96. Row (b) replaces the value of the deceased (in logarithm) from the estimated value of -7.32 to -10. Row (c) replaces the estimated state-specific death rates with a homogeneous value of 0.0164. Row (d) replaces the estimated state-specific recovery rates with a homogeneous value of 0.697.

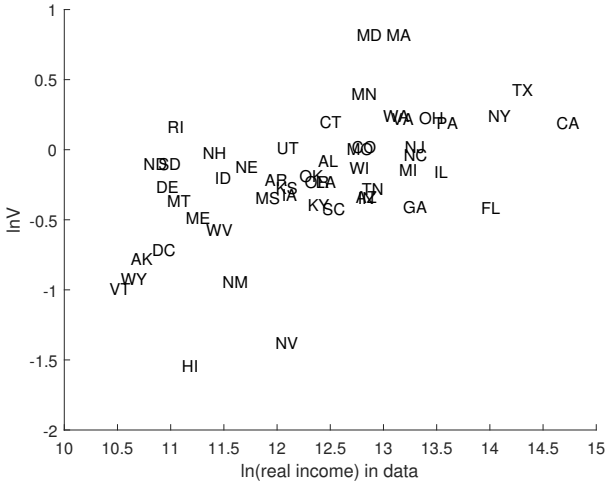
B.3 Additional Figures

Figure B.1: Histogram for Estimated High-Frequency Mobility Costs



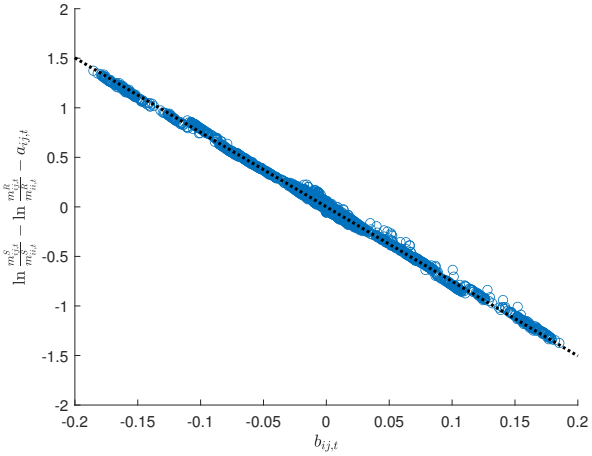
Notes. This figure plots the distribution of the estimated state-to-state mobility costs before the outbreak of the COVID-19 pandemic using the *PlaceIQ* data from January 2020. See Section 5.1 for the details on the estimation.

Figure B.2: Estimated Expected Utility and Real Income by State



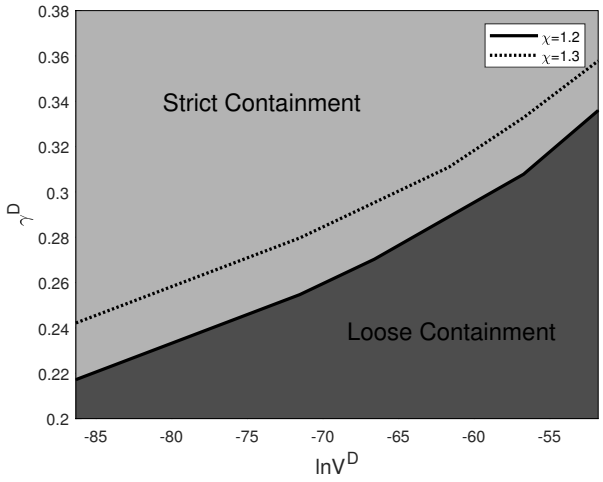
Notes. This figure plots the estimated value of staying in each state ($\ln(V)$) against each state's 2019 real income (in logarithm). See Section 5.1 for the details on the estimation.

Figure B.3: Estimation of the Value of the Deceased (V^D)



Notes. This figure represents the relationship between the estimated mobility differences of the susceptible and recovered individuals and the value of the deceased, according to equation (34). The vertical axis is the estimated value of the left-hand side of the equation: $\ln \frac{m^S_{ij,t}}{m^S_{ii,t}} - \ln \frac{m^R_{ij,t}}{m^R_{ii,t}} - a_{ij,t}$ while the horizontal axis is the estimated value of $b_{ij,t}$. Equation (34) implies that the slope of the fitted line of the scatter plots is $\ln(V^D)$, the logarithm of the value of the deceased.

Figure B.4: Optimal Policy for Cross-Region Mobility Control



Notes. This figure plots the estimated optimal policy for controlling mobility during a pandemic in the three-region economy specified in Section 4.2. The horizontal axis is the value of the deceased. The vertical axis is the death rate. The shaded plane is divided into two phases: "Strict Containment" is when the government prefers adopting mobility control policies as strictly as possible, and "Loose Containment" is as loosely as possible. The solid line divides the two phases when transmission rate $\chi = 1.2$ and the dashed line is when $\chi = 1.3$.

B.4 Relative Time Differences using the Dynamic Hat Algebra

Following Caliendo et al. (2019) to rewrite the value function of type S from equation (8) in terms of time differences, we find that:

$$\begin{aligned}
\dot{V}^S_{i,t+1} &= \frac{V^S_{i,t+1}}{V^S_{i,t}} = \frac{e^{u_i/\kappa}\sum_{j=1}^N (V^S_{j,t+2})^{\beta(1-\alpha_{j,t+2})}(V^I_{j,t+2})^{\beta\alpha_{j,t+2}}\mu^{-1}_{ij,t+1}}{e^{u_i/\kappa}\sum_{j=1}^N (V^S_{j,t+1})^{\beta(1-\alpha_{j,t+1})}(V^I_{j,t+1})^{\beta\alpha_{j,t+1}}\mu^{-1}_{ij,t}} \\
&= \sum_{k=1}^N \frac{(V^S_{k,t+1})^{\beta(1-\alpha_{k,t+1})}(V^I_{k,t+1})^{\beta\alpha_{k,t+1}}\mu^{-1}_{ik,t}}{\sum_{j=1}^N (V^S_{j,t+1})^{\beta(1-\alpha_{j,t+1})}(V^I_{j,t+1})^{\beta\alpha_{j,t+1}}\mu^{-1}_{ij,t}} \frac{(V^S_{k,t+2})^{\beta(1-\alpha_{k,t+2})}(V^I_{k,t+2})^{\beta\alpha_{k,t+2}}\mu^{-1}_{ik,t+1}}{(V^S_{k,t+1})^{\beta(1-\alpha_{k,t+1})}(V^I_{k,t+1})^{\beta\alpha_{k,t+1}}\mu^{-1}_{ik,t}} \\
&= \sum_{k=1}^N m^S_{ik,t} \frac{(V^S_{k,t+2})^{\beta(1-\alpha_{k,t+2})}(V^I_{k,t+2})^{\beta\alpha_{k,t+2}}\mu^{-1}_{ik,t+1}}{(V^S_{k,t+1})^{\beta(1-\alpha_{k,t+1})}(V^I_{k,t+1})^{\beta\alpha_{k,t+1}}\mu^{-1}_{ik,t}} \\
&= \sum_{k=1}^N m^S_{ik,t} \left(\dot{V}^S_{k,t+2}\right)^{\beta(1-\alpha_{k,t+2})} \left(\dot{V}^I_{k,t+2}\right)^{\beta\alpha_{k,t+2}} \left(\frac{V^S_{k,t+1}}{V^I_{k,t+1}}\right)^{\beta(\alpha_{k,t+1}-\alpha_{k,t+2})} \dot{\mu}^{-1}_{ik,t+1}.
\end{aligned}$$

B.5 The Normalized Hat Algebra Algorithm

First, we note that there will be no type I agent at the steady state. Therefore, the infection rate $\alpha = 0$. The steady state of the model is characterized by:

$$\begin{aligned}
V^S_{i,ss} &= \exp(\frac{u_i}{\kappa}) \sum_{j=1}^N \left(V^S_{j,ss}\right)^{\beta} (\mu_{ij,ss})^{-1}, \\
V^I_{i,ss} &= \exp(\frac{u_i}{\kappa}) \sum_{j=1}^N \left(V^I_{j,ss}\right)^{\beta(1-\gamma^R_j-\gamma^D_j)} \left(V^R_{j,ss}\right)^{\beta\gamma^R_j} (V^D)^{\beta\gamma^D_j} (\mu_{ij,ss})^{-1}, \\
V^R_{i,ss} &= \exp(\frac{u_i}{\kappa}) \sum_{j=1}^N \left(V^R_{j,ss}\right)^{\beta} (\mu_{ij,ss})^{-1}, \\
m^S_{ij,ss} &= \frac{\left(V^S_{j,ss}\right)^{\beta} (\mu_{ij,ss})^{-1}}{\sum_{k=1}^N \left(V^S_{k,ss}\right)^{\beta} (\mu_{ik,ss})^{-1}}, \\
m^I_{ij,ss} &= \frac{\left(V^I_{j,ss}\right)^{\beta\left(1-\gamma^R_j-\gamma^D_j\right)} \left(V^R_{j,ss}\right)^{\beta\gamma^R_j} (V^D)^{\beta\gamma^D_j} (\mu_{ij,ss})^{-1}}{\sum_{k=1}^N \left(V^I_{k,ss}\right)^{\beta(1-\gamma^R_k-\gamma^D_k)} \left(V^R_{k,ss}\right)^{\beta\gamma^R_k} (V^D)^{\beta\gamma^D_k} (\mu_{ik,ss})^{-1}}, \\
m^R_{ij,ss} &= \frac{\left(V^R_{j,ss}\right)^{\beta} (\mu_{ij,ss})^{-1}}{\sum_{k=1}^N \left(V^R_{k,ss}\right)^{\beta} (\mu_{ik,ss})^{-1}}.
\end{aligned}$$

The size of type S and R in each region satisfies

$$S_i = \sum_{j=1}^{N} S_j m_{ji,ss}^S,\ R_i = \sum_{j=1}^{N} R_j m_{ji,ss}^R.$$

Then for given $\kappa,\ \beta,\ \{u_i, \gamma_i^R, \gamma_i^D\}_{i=1}^N,\ \{\mu_{ij}\}_{N\times N}$, we can solve for the steady-state value function and mobility rate of S, I, and R using the equations above.

Denote $\widehat{x}_t = x_t/x_{ss}$ where x_{ss} is the steady-state value of x, following Proposition 7, the model can be rewritten as:

$$\widehat{V}_{i,t}^S = \sum_{j=1}^{N} m_{ij}^S \left(\widehat{V}_{j,t+1}^S\right)^{\beta(1-\alpha_{j,t+1})} \left(\widehat{V}_{j,t+1}^I\right)^{\beta\alpha_{j,t+1}} \left(\frac{V_j^I}{V_j^S}\right)^{\beta\alpha_{j,t+1}}, \tag{E.1}$$

$$\widehat{V}_{i,t}^I = \sum_{j=1}^{N} m_{ij}^I \left(\widehat{V}_{j,t+1}^I\right)^{\beta(1-\gamma_j^R-\gamma_j^D)} \left(\widehat{V}_{j,t+1}^R\right)^{\beta\gamma_j^R}, \tag{E.2}$$

$$\widehat{V}_{i,t}^R = \sum_{j=1}^{N} m_{ij}^R \left(\widehat{V}_{j,t+1}^R\right)^{\beta}, \tag{E.3}$$

$$m_{ij,t}^S = \frac{m_{ij}^S \left(\widehat{V}_{j,t+1}^S\right)^{\beta(1-\alpha_{j,t+1})} \left(\widehat{V}_{j,t+1}^I\right)^{\beta\alpha_{j,t+1}} \left(\frac{V_j^I}{V_j^S}\right)^{\beta\alpha_{j,t+1}}}{\sum_{k=1}^N m_{ik}^S \left(\widehat{V}_{k,t+1}^S\right)^{\beta\left(1-\alpha_{k,t+1}\right)} \left(\widehat{V}_{k,t+1}^I\right)^{\beta\alpha_{k,t+1}} \left(\frac{V_k^I}{V_k^S}\right)^{\beta\alpha_{k,t+1}}}, \tag{E.4}$$

$$m_{ij,t}^I = \frac{m_{ij}^I \left(\widehat{V}_{j,t+1}^I\right)^{\beta(1-\gamma_j^R-\gamma_j^D)} \left(\widehat{V}_{j,t+1}^R\right)^{\beta\gamma_j^R}}{\sum_{k=1}^N m_{ik}^I \left(\widehat{V}_{k,t+1}^I\right)^{\beta(1-\gamma_k^R-\gamma_k^D)} \left(\widehat{V}_{k,t+1}^R\right)^{\beta\gamma_k^R}}, \tag{E.5}$$

$$m_{ij,t}^R = \frac{m_{ij}^R \left(\widehat{V}_{j,t+1}^R\right)^{\beta}}{\sum_{k=1}^N m_{ik}^R \left(\widehat{V}_{k,t+1}^R\right)^{\beta}}. \tag{E.6}$$

The Algorithm

1. Solve the steady state V_i^S, V_i^I, V_i^R, m_{ij}^S, m_{ij}^I and m_{ij}^R.
2. For an exogenously given initial population distribution $\{L_0\}$, take a large enough number of periods, T.
3. Take initial guess $\left\{\widehat{V}_t^{S(0)}, \widehat{V}_t^{I(0)}, \widehat{V}_t^{R(0)}\right\}_{t=1}^T$.
4. For each iteration of $\left\{\widehat{V}_t^{S(k)}, \widehat{V}_t^{I(k)}, \widehat{V}_t^{R(k)}\right\}_{t=1}^T$:

(a) Given $\{S_{t-1}, I_{t-1}, R_{t-1}, D_{t-1}\}$, solve forward simultaneously for $\{S_t, I_t, R_t, D_t, \alpha_t\}$ from equations (10), (11), (12), (13), (E.4), (E.5), and (E.6).

(b) Use equations (E.1), (E.2), (E.3) to solve backward to get $\left\{\widehat{V}^{S(k)\prime}, \widehat{V}_t^{I(k)\prime}, \widehat{V}_t^{R(k)\prime}\right\}_{t=1}^{T}$

5. If $\left\{\widehat{V}_t^{S(k)\prime}, \widehat{V}_t^{I(k)\prime}, \widehat{V}_t^{R(k)\prime}\right\}$ is close to $\left\{\widehat{V}_t^{S(k)}, \widehat{V}_t^{I(k)}, \widehat{V}_t^{R(k)}\right\}$, finish. Otherwise, set next guess $\left\{\widehat{V}_t^{S(k+1)}, \widehat{V}_t^{I(k+1)}, \widehat{V}_t^{R(k+1)}\right\} = \left\{\widehat{V}_t^{S(k)\prime}, \widehat{V}_t^{I(k)\prime}, \widehat{V}_t^{R(k)\prime}\right\}$ and go back to step 4.

B.6 The Pandemic-free Model

In an economy with N regions, the mobility problem faced by agents when there is no pandemic is given by the following Bellman Equation:

$$U_{i,t}(\boldsymbol{\varepsilon_{i,t}}) = u_i + \max_j \left\{\beta E_t \left[U_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}})\right] - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t}\right\}.$$

If we define $V_{i,t} = \exp(E_{t-1}(U_{i,t})/\kappa)$ and $\mu_{ij,t} = \exp(\tilde{\mu}_{ij,t}/\kappa)$, following the proof of Proposition 1, it is easy to verify that

$$V_{i,t} = \exp\left(\frac{u_i}{\kappa}\right) \sum\nolimits_{j=1}^{N} \left(V_{j,t+1}\right)^{\beta} \left(\mu_{ij,t}\right)^{-1},$$

and the mobility rate from a region i to region j is given by

$$m_{ij,t} = \frac{\left(V_{j,t+1}\right)^{\beta} \left(\mu_{ij,t}\right)^{-1}}{\sum_{k=1}^{N} \left(V_{k,t+1}\right)^{\beta} \left(\mu_{ik,t}\right)^{-1}}.$$

B.7 Estimating Mobility Policy Elasticity

Firstly, from equation (21), which captures pre-pandemic mobility rates, we have:

$$\ln \frac{m_{ij}}{m_{ii}} + \ln \frac{m_{ji}}{m_{jj}} = -\ln \mu_{ij} - \ln \mu_{ji}. \tag{E.7}$$

Secondly, taking the difference between the equation (31) and (E.7) and substituting in equations (22) and (30) to get rid of the gravity variables in mobility costs, we find that

$$\ln \frac{m^g_{ij,t}}{m^g_{ii,t}} - \ln \frac{m_{ij}}{m_{ii}} + \ln \frac{m^g_{ji,t}}{m^g_{jj,t}} - \ln \frac{m_{ji}}{m_{jj}} = -b \cdot (Policy^{mob}_{i,t} + Policy^{mob}_{j,t}), \tag{E.8}$$

Thirdly, the aggregate mobility rate from a region i to j can be approximately decomposed as:

$$\begin{aligned} \ln m_{ij,t} &= \ln \frac{L_{ij,t}}{L_{i,t}} = \ln \left(\frac{S_{i,t} m_{ij}^S + I_{i,t} m_{ij,t}^I + R_{i,t} m_{ij,t}^R}{S_{i,t} + I_{i,t} + R_{i,t}} \right) \\ &\approx \tilde{S}_{i,t} \ln(m_{ij,t}^S) + \tilde{I}_{i,t} \ln(m_{ij,t}^I) + \tilde{R}_{i,t} \ln(m_{ij,t}^R), \end{aligned} \tag{E.9}$$

where $\tilde{g}_{i,t} = g_{i,t}/L_{i,t}$, $g \in \{S, I, R\}$ and $\tilde{S}_{i,t} + \tilde{I}_{i,t} + \tilde{R}_{i,t} = 1$. Similarly, the probability of staying in region i is given by

$$\begin{aligned} \ln m_{ii,t} &= \ln \frac{L_{ii,t}}{L_{i,t}} = \ln \left(\frac{S_{i,t} m_{ii}^S + I_{i,t} m_{ii,t}^I + R_{i,t} m_{ii,t}^R}{S_{i,t} + I_{i,t} + R_{i,t}} \right) \\ &\approx \tilde{S}_{i,t} \ln(m_{ii,t}^S) + \tilde{I}_{i,t} \ln(m_{ii,t}^I) + \tilde{R}_{i,t} \ln(m_{ii,t}^R). \end{aligned} \tag{E.10}$$

Therefore, taking the difference of (E.10) and (E.11), we have

$$\ln \frac{m_{ij,t}}{m_{ii,t}} \approx \tilde{S}_{i,t} \ln \frac{m_{ij,t}^S}{m_{ii,t}^S} + \tilde{I}_{i,t} \ln \frac{m_{ij,t}^I}{m_{ii,t}^I} + \tilde{R}_{i,t} \ln \frac{m_{ij,t}^R}{m_{ii,t}^R}. \tag{E.11}$$

Since we examine the initial phase of the COVID-19 pandemic, most of the population is susceptible to the virus, we assume $\tilde{S}_{i,t} \approx 1$, $\tilde{I}_{i,t} \approx 0$ and $\tilde{R}_{i,t} \approx 0$, equation (E.11) implies that

$$\ln \frac{m_{ij,t}}{m_{ii,t}} \approx \ln \frac{m_{ij,t}^S}{m_{ii,t}^S}. \tag{E.12}$$

Substitute the equation above into equation (E.8) for type S, we find that

$$\ln \frac{m_{ij,t}}{m_{ii,t}} - \ln \frac{m_{ij}}{m_{ii}} + \ln \frac{m_{ji,t}}{m_{jj,t}} - \ln \frac{m_{ji}}{m_{jj}} \approx -b \cdot (Policy_{i,t}^{mob} + Policy_{j,t}^{mob}). \tag{E.13}$$

B.8 Derivation of the Estimation Equation of V^D

First, from equation (9), the mobility rate of type S relative to R is given by

$$\ln \frac{m_{ij,t}^S}{m_{ii,t}^S} - \ln \frac{m_{ij,t}^R}{m_{ii,t}^R} = \beta \left[(1 - \alpha_{j,t+1}) \ln \frac{V_{j,t+1}^S}{V_{j,t+1}^R} - (1 - \alpha_{i,t+1}) \ln \frac{V_{i,t+1}^S}{V_{i,t+1}^R} + \alpha_{j,t+1} \ln \frac{V_{j,t+1}^I}{V_{j,t+1}^R} - \alpha_{i,t+1} \ln \frac{V_{i,t+1}^I}{V_{i,t+1}^R} \right] \tag{E.14}$$

Second, from equations (8) and (9), we can linearize the relative value equations as[37]

$$\ln \frac{V^I_{i,t}}{V^R_{i,t}} \approx \sum_j m^R_{ij,t} \left[\beta(1-\gamma^R-\gamma^D) \ln \frac{V^I_{j,t+1}}{V^R_{j,t+1}} + \beta\gamma^D \ln \frac{V^D}{V^R_{j,t+1}} \right] \tag{E.15}$$

$$\ln \frac{V^S_{i,t}}{V^R_{i,t}} \approx \sum_j m^R_{ij,t} \left[\beta(1-\alpha_{j,t+1}) \ln \frac{V^S_{j,t+1}}{V^R_{j,t+1}} + \beta\alpha_{j,t+1} \ln \frac{V^I_{j,t+1}}{V^R_{j,t+1}} \right] \tag{E.16}$$

We further assume that changes in the relative value function between two consecutive periods are sufficiently small, i.e., $\ln \frac{V^I_{i,t}}{V^R_{i,t}} \approx \ln \frac{V^I_{j,t+1}}{V^R_{j,t+1}}$ and $\ln \frac{V^S_{i,t}}{V^R_{i,t}} \approx \ln \frac{V^S_{j,t+1}}{V^R_{j,t+1}}$ for all j. Then equations (E.15) and (E.16) imply that

$$\ln \frac{V^I_{j,t+1}}{V^R_{j,t+1}} \approx a^R_{jt} + b^R_{jt} ln V^D \tag{E.17}$$

$$\ln \frac{V^S_{j,t+1}}{V^R_{j,t+1}} \approx a^S_{jt} + b^S_{jt} ln V^D. \tag{E.18}$$

while a^R_{jt} and b^R_{jt} are functions of β, γ^R, γ^D, $m^R_{ij,t}$, and $V^R_{j,t+1}$; a^S_{jt} and b^S_{jt} are functions of β, γ^R, γ^D, $m^R_{ij,t}$, $V^R_{j,t+1}$ and $\alpha_{j,t+1}$. Substitute equations (E.17) and (E.18) back to equation (E.14), we get:

$$\ln \frac{m^S_{ij,t}}{m^S_{ii,t}} - \ln \frac{m^R_{ij,t}}{m^R_{ii,t}} \approx a_{ij,t} + \ln V^D \cdot b_{ij,t}, \tag{E.19}$$

while $a_{ij,t}$ and $b_{ij,t}$ are functions of β, γ^R, γ^D, $m^R_{ij,t}$, $V^R_{j,t+1}$ and $\alpha_{j,t+1}$.

B.9 Extension: a Model with Vaccinations

We now generalize the baseline model to incorporate vaccinations. We assume that people who get vaccinated will be permanently immune, thus joining group R. We further assume that region i randomly vaccinate a fraction δ_i of the susceptible population in each period, so a susceptible person from region i moving to region j will have a chance of δ_j directly transferring to type R. Now the Bellman equation for group S people becomes:

$$\begin{aligned} U^S_{i,t}(\boldsymbol{\varepsilon_{i,t}}) = & u_i + \max_j \Big\{ \beta E_t \left[(1-\delta_j)(1-\alpha_{j,t+1}) U^S_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}}) + (1-\delta_j)\alpha_{j,t+1} U^I_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}}) \right. \\ & \left. + \delta_j U^R_{j,t+1}(\boldsymbol{\varepsilon_{j,t+1}}) \right] - \widetilde{\mu}_{ij,t} + \varepsilon_{ij,t} \Big\}. \end{aligned}$$

[37] The approximations are quite accurate. We compare the estimated value of V^R_{it} from linear approximation and non-linear exact solution using the parameters we estimate. Their difference is less than 1%.

As for the demographic dynamic, a fraction δ_i of type S will join type R:

$$\overline{S}_{i,t} = (1-\delta_i)\sum_{j=1}^{N} S_{j,t-1} m^S_{ji,t-1},\ \overline{R}_{i,t} = \sum_{j=1}^{N} R_{j,t-1} m^R_{ji,t-1} + \delta_i \sum_{j=1}^{N} S_{j,t-1} m^S_{ji,t-1},$$

therefore, vaccination can slow down the spread of pandemics directly. If $\delta_i = 0$, the model collapses to our benchmark. We can show that the value function of type S satisfies

$$V^S_{i,t} = \exp\left(\frac{u_i}{\kappa}\right) \sum_{j=1}^{N} \left(V^S_{j,t+1}\right)^{\beta(1-\delta_j)(1-\alpha_{j,t+1})} \left(V^I_{j,t+1}\right)^{\beta(1-\delta_j)\alpha_{j,t+1}} \left(V^R_{j,t+1}\right)^{\beta\delta_j} \left(\mu_{ij,t}\right)^{-1},$$

while the value functions of types I and R remain the same as Proposition 1. The mobility rate of type S is given by

$$m^S_{ij,t} = \frac{\left(V^S_{j,t+1}\right)^{\beta(1-\delta_j)(1-\alpha_{j,t+1})} \left(V^I_{j,t+1}\right)^{\beta(1-\delta_j)\alpha_{j,t+1}} \left(V^R_{j,t+1}\right)^{\beta\delta_j} \left(\mu_{ij,t}\right)^{-1}}{\sum_{k=1}^{N} \left(V^S_{k,t+1}\right)^{\beta(1-\delta_k)(1-\alpha_{k,t+1})} \left(V^I_{k,t+1}\right)^{\beta(1-\delta_k)\alpha_{k,t+1}} \left(V^R_{k,t+1}\right)^{\beta\delta_k} \left(\mu_{ik,t}\right)^{-1}},$$

while the mobility rates of types I and R remain the same as Proposition 2.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1980	Alan Manning Graham Mazeine	Should I stay or should I go? Return migration from the United States
1979	Rosa Sanchis-Guarner José Montalbán Felix Weinhardt	Home broadband and human capital formation
1978	Luca Macedoni John Morrow Vladimir Tyazhelnikov	Firms in product space: Adoption, growth and competition
1977	Gonzalo Nunez-Chaim Henry G. Overman Capucine Riom	Does subsidising business advice improve firm performance? Evidence from a large RCT
1976	Julian Alves Bruno Serra Jason Greenberg Yaxin Guo Ravija Harjai John Van Reenen	Labour market power: New evidence on Non-Compete Agreements and the effects of M&A in the UK
1975	Virginia Minni	Global managers, local workers: Wage setting inside a multinational firm
1974	Jonathan Colmer Suvy Qin John Voorheis Reed Walker	The changing nature of pollution, income and environmental inequality in the United States
1973	Italo Colantone Gianmarco I.P. Ottaviano Piero Stanig	The social footprint of globalization: Towards the introduction of strategic industries in quantitative trade models
1972	Viet Nguyen-Tien Robert J.R. Elliott Eric Strobl Chengyu Zhang	Estimating the longevity of electric vehicles: What do 300 million MOT test results tell us?

1971	Michael Amior Jan Stuhler	Immigration, monopsony and the distribution of firm pay
1970	Jonathan Colmer David Lagakos Martin Shu	Is the electricity sector a weak link in development?
1969	Gaia Dossi Marta Morando	Political ideology and innovation
1968	Natalie Irmert Jan Bietenbeck Linn Mattisson Felix Weinhardt	Autonomous schools, achievement and segregation
1967	Stephen B. Billings Adam Soliman	The erosion of homeownership and minority wealth
1966	Pawel Bukowski Pawel Chrostek Filip Novokmet Marek Skawiński	Income inequality in the 21st century Poland
1965	Benny Kleinman Ernest Liu Stephen J. Redding Motohiro Yogo	Neoclassical growth in an interdependent world
1964	Hanwei Huang Gianmarco I.P. Ottaviano	Rethinking revealed comparative advantage with micro and macro data
1963	Natalie Chen Dennis Novy Carlo Perroni Horng Chern Wong	Urban-biased structural change
1962	Robin Kaiji Gong Yao Amber Li Kalina Manova Stephen Teng Sun	Tickets to the global market: First US patent awards and Chinese firm exports