informs
https://pubsonline.informs.org/journal/ijds

INFORMS JOURNAL ON DATA SCIENCE
Vol. 3, No. 1, April–June 2024, pp. 6–27
ISSN 2694-4022 (print), ISSN 2694-4030 (online)

The Interplay Between Individual Mobility, Health Risk, and Economic Choice: A Holistic Model for COVID-19 Policy Intervention

Zihao Yang,[a] Ramayya Krishnan,[a] Beibei Li,[a,*]

[a] Information Systems and Management, Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
*Corresponding author

Contact: zihaoyan@andrew.cmu.edu (ZY); rk2x@cmu.edu, https://orcid.org/0000-0001-9935-2468 (RK); beibeili@andrew.cmu.edu, https://orcid.org/0000-0001-5466-7925 (BL)

Received: July 6, 2023
Revised: January 20, 2024; April 10, 2024
Accepted: April 20, 2024
Published Online in Articles in Advance: May 14, 2024

https://doi.org/10.1287/ijds.2023.0013

Abstract. This paper was motivated by the need to simultaneously address two competing policy objectives during the course of the COVID pandemic: namely, the public health objective, which required people to be less mobile, and the economic objective, which aimed to ensure that the economy was not adversely affected by the constraints imposed by the first objective. To realize these objectives, we developed a data-informed approach to model human mobility, health risk, and economic activity jointly. This approach computes equilibrium between epidemic models of public health and economic activity under policy interventions that could be used to change people's mobility behavior. Our approach is distinctive in its capacity to assemble proprietary data sets from public and private sectors at the individual and the zip code levels, which heretofore had not been used together. These data enabled customization of the population-level epidemic models widely used in public health (e.g., the SIR model) with individual-level data traces of mobility behaviors for assessment of public health risks. The outputs of the proposed model enabled parameterization of economic choice models of individuals' economic decision-making. Various policy interventions and their capacities to shift the equilibrium between economic activity and public health were investigated in this study. Whereas the data-informed joint modeling approach was developed and tested in the pandemic context, it is generalizable for the evaluation of any counterfactual policy interventions.

History: Olivia R. Liu Sheng and W. Nick Street served as senior editors for this article.
Data Ethics & Reproducibility Note: The code capsule is available on Code Ocean at https://doi.org/10.24433/CO.4390192.v1 and in the e-Companion to this article (available at https://doi.org/10.1287/ijds.2023.0013).

Keywords: mobility analytics • pandemic policy design • economic choice • consumer location data • healthcare risk • epidemic modeling • data-driven policy making

1. Introduction

The pandemic (COVID-19) caused global havoc to both public health and the economy. From early 2020, governments adopted various policies in response to COVID-19 (Hale et al. 2020). For example, general policies included international travel bans, mask mandates, and stay-at-home requirements. These broad measures aimed to limit the spread of the virus without focusing on specific sectors or populations. Targeted policies were also implemented to address specific challenges. Specifically, supply-side policies, such as restrictions on internal movement and targeted closures, were designed to control the spread of the virus within particular regions or industries (Birge et al. 2022). Demand-side policies, such as stimulus checks, aimed to bolster the economy and support those affected by the pandemic (Kim and Lee 2021). Although these policies had an impact on mitigating healthcare risks whereas supporting the economy, several challenges remained.

First, individuals' health risks and their responses to pandemic policies are often heterogeneous. However, existing policy designs were largely homogeneous and did not account for individuals' socioeconomic factors (e.g., income, age) or behavioral motivations. For instance, among the 17 mainstream government policies documented in Hale et al. (2020), 16 were homogeneous in design; only 1 policy considered differentiating a specific socioeconomic group based on income support.

Second, existing pandemic policy designs focused mostly on reducing healthcare risk, with limited understanding of its continuing impact on individuals and the local economy. Besides, existing policies tend to take either a supply-side (e.g., targeted or nontargeted closures) or demand-side (e.g., economic subsidy) approach. However, all these critical factors interact in a complex system. For example, individuals assess the health risks of a destination before deciding to visit. The health risk of a location depends on how many people have visited it in the past. The prevailing COVID-19 policy, the extent of compliance with the policy, socioeconomic factors, the economic preferences of individuals, and the types of places being visited all interact in complex ways. The underlying mechanism of this complex system has to be modeled prior to developing an understanding of how people respond to different policy designs. This can be highly challenging because it requires a deep understanding of how exactly individuals respond to the spread of COVID-19 (e.g., decisions to visit certain places versus stay at home) and the underlying social and economic drivers of such behaviors.

Third, there is a lack of understanding of how existing policies can alleviate social and economic inequality during pandemics. Although our society as a whole was affected by the epidemic, the infection rate disparity among different socio-demographic groups is well documented (Chowkwanyun and Reed 2020, Garg et al. 2020, Hardy and Logan 2020, Pareek et al. 2020, Van Dorn et al. 2020, Yancy 2020). Because human mobility is fundamental to the spread of infectious diseases (Balcan et al. 2010, Poletto et al. 2013, Courtemanche et al. 2021), one way to explain that disparity is to study the mobility pattern in different social groups under COVID-19. For example, Chang et al. (2021) found that disadvantaged groups were not able to reduce their mobility as sharply as privileged groups, which might have contributed to the higher infection rate. However, *what causes the difference in mobility patterns between different groups? What is the decision-making process that determines whether to stay at home or go out during an epidemic? What can policymakers do to reduce infection rates whereas supporting social welfare?*

To answer these questions, we proposed a holistic model to jointly model healthcare risk, economic activity, and human mobility using state administrative data on COVID-19 infections along with two unique large-scale private sector data sets that capture individual-level mobility data and individual-level bank transaction data from a major financial institution. The hypothesis is that health risks and economic activity are crucial factors in mobility decisions and, consequently, in the spread of the epidemic. With economic activity as a moderator, we are able to explain the difference in mobility patterns between different income groups. Based on an understanding of the interplay between human mobility, health risk, and economic activity, we are able to assess and recommend public policies to promote social welfare.

More specifically, we modeled this holistic process as three interactive components: (1) the effect of human mobility on health risk, (2) the effect of health risk and economic activity on human mobility, and (3) the equilibrium.

First, to capture the impact of human mobility on COVID-19 spread, we introduced an individual-level susceptible-infected-recovered model (individual SIR model) that maps individual human mobility and interactions with specific points-of-interest (POI) categories (e.g., restaurants, grocery stores, gas stations) to the spread of COVID-19. Prior research mostly used aggregated mobility data (Badr et al. 2020, Jia et al. 2020, Lai et al. 2020), synthetic mobility data (Block et al. 2020, Duque et al. 2020), or pure mobility data without an embedded epidemiological model (Benzell et al. 2020, Block et al. 2020, Hsiang et al. 2020). In contrast, we build a dynamic network based on the SIR epidemiological model, using real-life mobile phone geo-location data (tracking more than 10 billion movement records of 10 million unique individuals in Pennsylvania) as input and zip-code-level measures of COVID-19 daily cases as output.

The dynamic model we propose is among the first to incorporate individual-level movement trajectory into the simulation of the spread of an epidemic. It enables us to account for not only the general spread of COVID-19 as measured by daily zip code–specific case numbers and death rates but also, more importantly, POI category–specific health risk as captured by the number, source, and risk of individual visitors being colocated. This data-driven approach allows us to predict the spread of COVID-19 more accurately, highlight locations that can be "super-spreaders" of COVID-19 in real time, and reveal the health risk disparities among different socio-demographic groups. Second, to estimate the impact of COVID-19 health risks on human mobility, we took an economic perspective in modeling individuals' mobility choices. Specifically, we modeled a consumer's choice to visit a place in person (e.g., shopping offline) or not (e.g., shopping online) as a function of health risk, economic constraints (e.g., income, travel distance, any economic subsidy such as delivery fees waivers, etc.) and other factors such as temporal effects (e.g., weekend versus weekday). In addition, we focused primarily on individuals' economic activity for this model because it was a major driver of consumer mobility during COVID-19 (Bonaccorsi et al. 2020, de Palma et al. 2022). We validated this binary choice model with a unique large-scale credit card transaction data set from a major regional bank, containing detailed zip-code-level transaction information for Pennsylvania from February 2020 to May 2021, comprising 12.6 million records and accounting for USD 840 million in total spending per month.

Finally, a negative feedback loop of the mutual effects of COVID-19 spread and human activity: A decrease in infection rate will lead people into more in-store purchasing activity and, in turn, increase the infection rate. The equilibrium established on the negative feedback loop enabled us to identify changes in infection rate and in-store purchase probability. Thus, we could simulate various counterfactual policy interventions (e.g., distribution of delivery coupons, enforcement of mask mandates, or category-specific lockdowns) and evaluate the corresponding effectiveness by quantitatively changing factors in the equilibrium.

Our model demonstrated significantly higher performance compared with a number of baselines. We tested our individual SIR model via out-of-sample prediction, and it outperformed the mobility-based SIR model lacking individual mobility data by 21% and the baseline SIR model without any mobility data by 60% in predicting zip code–level COVID case numbers. Furthermore, our model-estimated industry-category–specific risk coefficients were shown to be highly consistent with the Risk Assessment Chart (Texas Medical Association 2020) from the Texas Medical Association (i.e., with a small Kendall Tau distance of $3/21 = 0.14$ between the two).

We found from the individual SIR model that household size and age positively correlate with COVID-19 transmission probability, whereas income negatively correlates with COVID-19 transmission probability. Interestingly, we did not find that low-income individuals visit locations that belong to industry categories of higher health risk; instead, we found that they tend to visit more crowded locations under the same industry category than the general public. For example, both the high-income and low-income groups visit grocery stores, but the latter tend to visit those branches that have a higher density of colocated visitors, which condition contributes to a higher health risk. This is likely because low-income groups have mobility constraints and, hence, might have limited location choices in the consideration set for economic activity. We also found that among the low-income group, more multigenerational households contribute to higher infection rates.

The economic choice model showed that exposure risk of COVID and the average distance from a POI are significantly positively correlated with online purchase probability, whereas delivery fee and weekend effects are significantly negatively correlated with online purchase probability. In addition, our findings are consistent with prior literature demonstrating a significant and positive relationship between income level and online shopping probability (Farag et al. 2006, Bell and Song 2007).

By linking the mobility and economic choice models, we were able to find an equilibrium between health risk and human mobility. Moreover, the interaction between the two models allowed us to test how such equilibrium would change under different pandemic policies. Specifically, we tested three counterfactual settings: giving out delivery coupons, enforcing mask mandates, and announcing categorical lockdowns. We found that giving out delivery coupons will reduce the infection rate and have greater effects on low-income groups. Interestingly, we also found that whereas mask mandates, in theory, will curb virus spread, loose enforcement may inadvertently increase the overall infection rate. In the categorical lockdown setting, we found that a specific lockdown in Pennsylvania, encompassing restaurants, shopping malls, and grocery stores, led to a uniform reduction in infection rates across all zip codes, but its effectiveness was slightly diminished when taking into account a shift to online shopping.

Previous research mostly focused on mobility's impact on disease spread by building a prediction model of COVID-19 with case numbers (Putra and Mutamar 2019, Cooper et al. 2020, Salgotra et al. 2020), or with aggregated mobility data (Kraemer et al. 2020, Nouvellet et al. 2021). Other researchers investigated the mobility changes in response to COVID-19 (Abu-Rayash and Dincer 2020, Warren and Skillman 2020). The novelty of the present work lies in the following three areas.

First, it is the first study to model epidemic spreading and individual movement trajectories instead of relying on fitting aggregated data sets. The individual SIR Model was shown empirically to be more capable of predicting zip code–level COVID-19 cases than the population-level SIR model. It is also structurally more flexible, allowing the integration of more sociodemographic information and accounting for the density, path, and associated risk of colocated individuals in real time in estimating POI category–specific healthcare risk. Such an approach also enabled us to better capture the individual-level data generation process, thus enabling a deeper understanding of the underlying mechanisms of the observed human mobility patterns and associated health risks.

Second, the literature in this area has mostly treated human mobility as given or assumed to follow an exogenous stochastic process. Instead, we aimed to endogenize human mobility decisions by taking an economic perspective in modeling human mobility behavior as a utility maximization process that is subject to both health risk and economic constraints. By doing so, we could further capture the individual-level data generation process of human mobility.

Third, to date, there is a lack of work that combines healthcare risk data, consumer-level economic consumption data, and mobility data to capture the feedback loop between human mobility patterns and the spread of epidemics. In this study, we leveraged both epidemiological modeling and economic theories and established the equilibrium to understand the dynamic interactions between individual economic activity, mobility, and health risk in a holistic fashion.

2. Related Literature

Since its breakout in 2020, COVID-19 has become a hot topic in epidemiology, sociology, economics, machine learning, and data science research. Our work is closely related to the literature on the prediction of COVID-19 cases and the analysis of the economic and public health effects of COVID-19.

2.1. Epidemiology Models: Analysis and Prediction of COVID-19 Spread

In the mathematical modeling context, there is much research on different aspects of the spread of COVID-19 using the SIR model, which is one of our bases for COVID-19 case prediction in Section 4. In Cooper et al. (2020), the authors investigate the spread of the disease in different communities over time. After monitoring the diversity in significant parameters, the authors find that suitable early infection rates can manage COVID-19 under control. In Chen et al. (2020), the authors propose a time-dependent SIR model, the parameters of which can change over time, and claim that the model can be adapted to disease control policy entailing lockdowns and mask mandates. In Calafiore et al. (2020), the authors take the initial number of susceptible people and the unknown number of infected people into account in their model.

These papers show that the SIR model as an epidemiology model offers great utility for the prediction of COVID-19 spread and that, as a framework incorporating various features, it has great potential. However, the literature is based on population-level mobility data, not on individual level (Goel and Sharma 2020, Paoluzzi et al. 2021). The lack of integration of mobility data on a more granular level has several shortcomings.

First, the predictions of epidemiology models with high-level mobility data lack power. The population mobility data used in this research (Calafiore et al. 2020, Cooper et al. 2020) usually investigate only an aggregated transport flow from one geographical area (zip code or county) to another. That is, they do not take into account the latent variable of the socio-geographical background of individuals and their travel behaviors as influenced or determined by that background. As such, the prediction often overestimates the risk to people with lower travel frequency or in lower-risk areas and underestimates the risk to those who travel more frequently and to higher-risk areas. Moreover, the predictions of those models are often misleading. Because of the aggregated nature of population-level mobility data, research using it will inevitably treat mobility as a flow of population instead of a population of trajectories and thus will go on to attribute mobility to the wrong people, neglecting the path and associated risk of colocated individuals in real time, or omitting the order of a travel trajectory. For these reasons, integrating individual travel trajectories could reinstate the validity and boost the accuracy of disease-spread prediction.

One possible limitation of the individual SIR model may come from its inherently higher computational requirements. This increased demand is attributed to the model's finer granularity of data representation. Consequently, this characteristic may render the individual-based SIR model less suited to contexts where computational resources are constrained, such as in less developed regions or countries, or when real-time predictive capabilities are needed.

2.2. Socioeconomic Impact of COVID-19

The COVID-19 pandemic, since its outbreak in late 2019, has exerted significant socioeconomic impacts on a global scale (Nicola et al. 2020). These impacts have been diverse and far-reaching, affecting almost every aspect of society, from health and education to employment and the economy at large.

The economic impacts have been particularly devastating. The pandemic has led to an unprecedented global economic downturn, characterized by widespread business closures and soaring unemployment rates (Coibion et al. 2020). The International Monetary Fund (IMF) reported a contraction of the global economy by an estimated 3.5% in 2020 alone (International Monetary Fund 2020). This contraction was not restricted to particular sectors or regions; rather, it was a widespread phenomenon affecting both developed and developing economies. Despite attempts at economic recovery through various fiscal stimuli, numerous countries continue to face substantial challenges in restoring their economies to prepandemic levels (International Monetary Fund 2021).

Beyond the economic consequences, the pandemic has exacerbated social inequalities. COVID-19 has hit marginalized communities the hardest, intensifying pre-existing inequalities, such as education inequality (Van Dorn et al. 2020), gender inequality (Alon et al. 2020), and racial inequality (Sood and Sood 2021). Particularly, the impact of the pandemic on low-income families has been devastating, as these families are less able to weather the financial and health shocks brought about by the crisis (Adams-Prassl et al. 2020). Disparities in access to resources, including healthcare and online education, have also become more evident during the pandemic.

In addition, the pandemic has brought about a profound change in people's everyday lives. With lockdowns and social distancing measures being enforced worldwide, individuals and communities have experienced substantial mental health effects. Rising levels of stress, anxiety, and depression have been reported across different populations, underscoring the severe psychological impacts of the pandemic (Xiong et al. 2020).

One crucial area that the COVID-19 pandemic has dramatically affected is consumer behavior, specifically in the domain of retail and shopping; as a result of social distancing measures and lockdowns, traditional brick-and-mortar retail has been greatly impacted, leading to a significant shift toward e-commerce (Roggeveen and

Sethuraman 2020, Sayyida et al. 2021). The inability to shop in person, coupled with an increased amount of time spent at home, has accelerated the adoption of online shopping platforms across different demographics.

To comprehensively evaluate the transformation of consumer behavior under the influence of the COVID-19 pandemic, it is imperative that we critically examine the ramifications of this global crisis on the dynamics of online shopping. Koch et al. (2020) report that the economic situation can affect consumer's motivation in online shopping. Sharma and Jhamb (2020) use daily consumable sales data from an Indian e-commerce website as evidence to show that the losses and downshift caused by COVID-19 are not negligible. Nguyen et al. (2020) collect survey-based data from Vietnamese consumers to show that hedonic motivation prevails in online book purchasing. Overall, online shopping faces several fundamental obstacles in persuading more consumers to embrace it. In Huang and Oppewal (2006), the authors show that delivery fee is not as important as travel time to the grocery store when deciding whether to purchase online. Specifically, 15 minutes in traveling time has a heavier weight than a delivery fee of £5. Le et al. (2022) established the connection between online shopping and travel demand. Shi et al. (2020) use an interview-based method in China to show that online consumers are more likely to have longer travel distances and that therefore the lockdown might not be as useful as policymakers had thought. An interesting related study (Mantin and Koo 2010) of the weekend effect on airline prices shows a strong effect in terms of the dispersion of airfares. This motivated us to take the weekend effect on online shopping into consideration.

2.3. Consumer Choice Between Online vs. In-Store Shopping

Our work is related to literature that has examined consumers' decision making between online and in-store shopping. For example, Mosteller et al. (2014) propose a model of consumers' observation of the verbal fluency of online information's effect on the decision to purchase online. Also, in Schmid et al. (2016), the authors use a latent variable model to explain the online shopping convenience and socioeconomic characteristics' effect on the choice of online or in-store shopping. In Hsiao (2009), the authors conduct quantitative research to show that the value of delivery time for an online good is around \$0.53 per day. In Etminani-Ghasrodashti and Hamidi (2020), research conducted in Iran shows that building environment, store attributes, and consumers' lifestyles are the main factors affecting consumers' in-store purchasing choices.

However, consumer shopping behavior has changed significantly under COVID-19, for example, in the transition from offline to online shopping, avoidance of high-risk areas, and other aspects. Previous work in this area has not taken into account the interaction between consumer economic decisions and human mobility in the context of the pandemic. Notably, as higher-risk areas and merchant categories change over time, the utility of avoiding participation in such activity will vary.

3. Data and Description

3.1. Data

In this paper, we conducted analyses on a combination of five large unique data sets obtained from a variety of sources in the public and private sectors. We refer to them henceforth as *bank data set*, *mobility data set*, *POI data set*, *case data set*, and *survey data set*. We first introduce each of them in this section. Then, in the next section, we will discuss how we make the linkage among them.

3.1.1. Bank Data Set. We obtained bank data set (Table 1) from a major bank operating in 27 states in the United States. The data set includes consumer credit card spending information from a subsidiary of the bank. The data set is composed of transaction details, including transaction date, amount, type, means, and industry category.

The transaction date is described by the following three attributes: date of postage, date of completion, and date of transaction. The date of postage is the date on which the transaction gets posted in the banking system; the date of completion is the date that the transaction is completed and the bank wires money to the merchant; the date of the transaction is the date when

Table 1. Variable Description of Bank Data Set

Variable name	Variable description	Mean	Standard deviation
TransDate	Date of the transaction	—	—
TransAmt	Amount of money in transaction	293.48	635.73
TransCate	Industry-category of the transaction	—	—
TransMns	Means of the transaction (online or offline)	—	—
Income	Median income of consumer's CBG	90,856	9,046
Dist	Average distance to merchant	1.51	0.55

Table 2. Variable Description of Mobility Data Set

Variable category	Level	Variable name	Variable description
Location	Geographic	*Lat*	Latitude of recorded location
	Geographic	*Lon*	Latitude of recorded location
Individual	Person	*HomeAdr*	Home address of individual
	Person	*IDFA*	Identifier for advertising
POI	CBG	*PoiCbg*	CBG of the POI
	Category	*PoiCate*	Industry-category of the POI

the actual purchase happens. We selected the date of transaction, denoted as `TransDate`, for our purposes.

The transaction amount is the total amount of money transferred from the consumer to the merchant and is denoted as `TransAmt`.

The transaction type is used by the bank to identify and manage payments. The transaction types include Visa Authorization Adjustments, Case Authorization Adjustments, Returns, Sales, and others. We focused on sales data in our study.

The means of transaction, denoted as `TransMns`, is categorized into nine classes: mail/phone transaction, electronic transaction, face-to-face transaction, and others. Because migration between online and offline transactions is a matter of interest in this paper, we aggregated the nine categories into two classes: online transactions and offline transactions. Each transaction record has an identifier for the means of transaction, enabling us to identify each transaction as offline or online. Each record of the transaction has an identifier for the means of transaction. Therefore, we can identify each transaction as offline or online.

The industry category, denoted as `TransCate`, is a classification of merchants. Identified by the Standard Industrial Classification (SIC) code, merchants are classified into more than 10,000 categories, ranging from grocery stores to airlines. It was of great importance to our research to learn what the consumer spent money on and how different the proportions of online transactions were among the different industry categories.

For this study, our sample from the bank data set contains detailed spending information from all consumers in Pennsylvania dating from February 2020 to May 2021, with around 12.6 million individual records, $840 million USD total spending per month, of which around 90.0% number of records and total spending amount are identified as sales. 36.3% of the 12.6 million monthly transactions are online transactions, which accounts for 47.6% of the $840 million USD monthly total spending. Of the eight means of online transaction, channel-encrypted electronic commerce transaction is the most common way, contributing $260 million USD spending (65.0% of all online transactions) and 3.2 million records (71.6% of all online transactions) per month. By industry level, grocery stores, restaurants, and gas stations are the top three contributors, accounting for 12.0%, 10.8%, and 3.67% in total transaction amount, respectively.

Online spending changes over time and varies within industry categories, which will be discussed in Section 3.3.

3.1.2. Mobility Data Set. The *mobility data set* is a mobility information data set (Table 2) sourced from an international geo-location data provider. The data set is collected by a Software Development Kit (SDK), a module on which mobile phone applications are built. When mobile phone users agree to the terms and conditions of the application and start to use it, the location tracker embedded in the SDK will make a request to the mobile phone's global positioning system (GPS) and send it to the data provider the user' location at certain regular time intervals. The data provider's SDK is available on both Android and iOS.

Each individual record in the *mobility data set* contains a unique identifier for each mobile phone, which does not change over time. For simplicity, and given that the data are anonymous and advertiser ID is the only way to identify individuals, we recorded the same identifier as the records of the same person and recorded different identifiers as the records of different people.

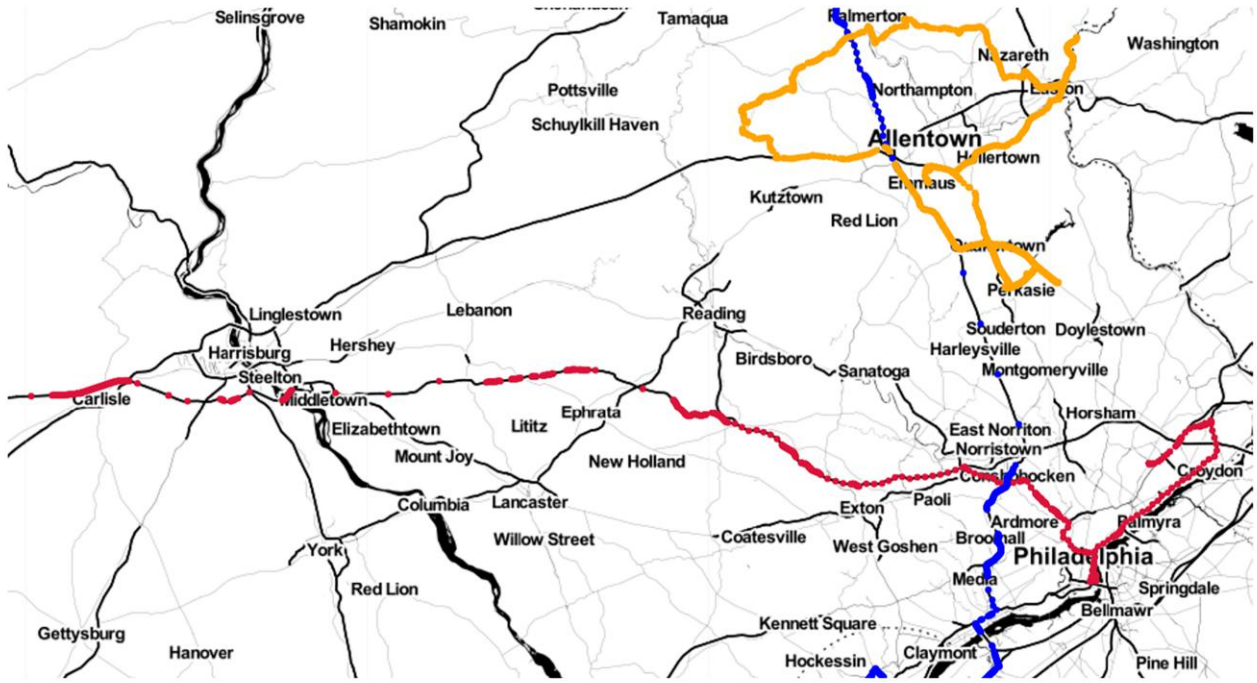
The mobility data set is packed into days, but there is no further timestamp stating the specific time when the record is tracked. Therefore, we conducted the epidemiology analysis in 4 on a daily level.

The geolocation of the user, which is where the user is visiting, is represented in mobility data set by latitude and longitude, and the error is under 10 meters.

We obtained mobility data set from January 2020 to November 2020, containing around 10 billion records in Pennsylvania and more than 300 billion records in the United States. For the purposes of this study, we focused on all individuals primarily residing in the state of Pennsylvania.

3.1.3. Other Data Sets. The POI data set is provided by SafeGraph, a global data provider. The data set contains around 170,000 points of interest in Pennsylvania, with their location (latitude and longitude) and industry (SIC code).

Figure 1. (Color online) Movement Trajectories Overlay



Note. Dots show records of three sample individuals in *mobility data set* on March 30, 2020.

We mapped each visit in mobility data set to points of interest in POI data set and discovered the industry category each visit is for via the SIC code in POI data set. Figure 1 is an illustration of movement trajectory in the data set.

The case data set is a proprietary data set provided by Pennsylvania's Department of Health. It contains every day's case number in each zip code in Pennsylvania, dating from May 2020 to May 2021.

Finally, the survey data set is from the COVID-19 Trends and Impact Survey conducted by the CMU Delphi Research Group via a partnership with Facebook (Salomon et al. 2021). Respondents were questioned about COVID-like symptoms, their behavior (such as social withdrawal), mental health, and the economic and health effects of the pandemic. The reported mask-wearing level was of particular interest to us, as we incorporated the attributes into both the epidemic spread model and the economic choice model. The survey ran from April 6, 2020, to June 25, 2022, and generated about 40,000 to 50,000 monthly responses in Pennsylvania. The attributes in which we were most interested were the self-reported mask-wearing level and the perceived mask-wearing level.

3.2. Overview of Data Structures

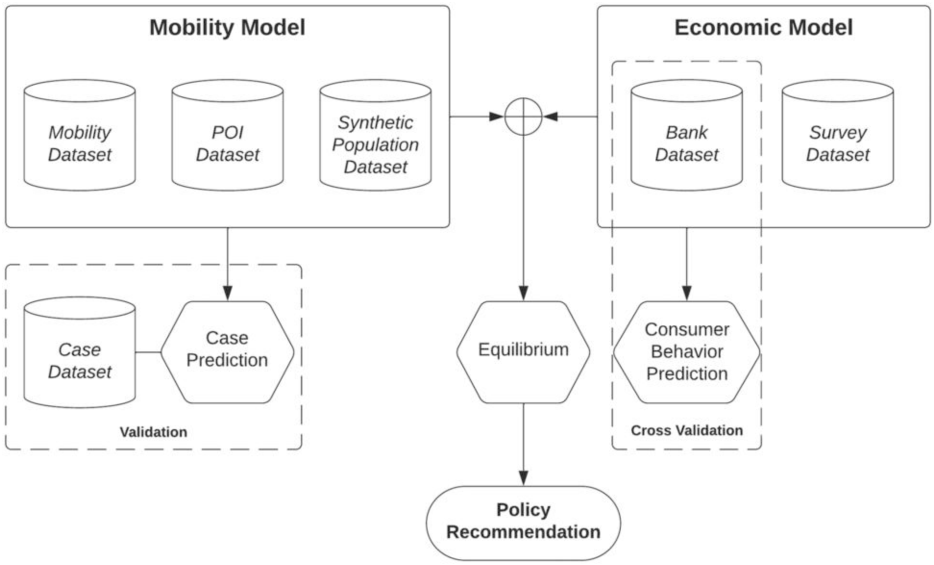
To understand the spread of COVID-19 and infection factors, we engage in the use of the mobility data set and POI data set.[1] These data sets are instrumental in the design of the Mobility Model, a predictive construct for COVID-19 transmission estimations. The Mobility Model was validated based on daily COVID-19 case numbers reported at the zip code level obtained via the case data set. Instead, we observed COVID-19 case numbers at the zip code level. Therefore, we modeled the mobility and infection processes at the individual level and then calibrated the model at the zip code level by aggregating the model-predicted probabilities for all individuals in a given zip code. We discuss further details of the Mobility Model in Section 4.

Meanwhile, in an effort to forecast consumer behaviors, we used the bank data set and survey data set. These data sets feed into the Economic Model, providing a dynamic representation of consumer activities and economic responses in the face of the pandemic. By synergizing the Mobility Model and the Economic Model, we constructed equilibria under counterfactual settings. We discuss the economic model and the equilibrium in Sections 5 and 6, respectively.

In summary, we illustrate an overview of the data and analysis structure in Figure 2.

Although one major advantage of our work is its leveraging of a wide variety of unique data sets from both the public and private sectors, some data were recorded at different granularities than others. We provide the time granularity and level of observation of each data set in Table 3. In our main analysis, we opted to harmonize the analysis at the granularity of the day and zip code levels. The reason was manifold as detailed here.

Figure 2. Overview of Data and Analysis Structure



Daily granularity permits the analysis of trends and changes on a day-to-day basis, revealing patterns or anomalies that might be obscured at a coarser time resolution. In the context of a pandemic, where situations evolve rapidly, daily data can be vital in recognizing emerging trends or shifts in the progression of the disease, thereby allowing for more responsive policy decisions and interventions.

The use of zip codes provides a balance between detail and privacy. It offers a localized perspective that is more precise than that of larger geographic units (like a city or state) but less so than individual addresses, thus preserving some level of privacy. This granularity can reveal localized trends and disparities, enabling targeted interventions in specific areas showing higher rates of infection, economic distress, or other indicators of interest. Furthermore, incorporating data at the census block group (CBG) level would prove less effective and potentially infeasible.

Besides, conducting research at the zip code level allows for a community-focused understanding of the situation. Zip codes can be representative of specific communities or neighborhoods, each with distinct demographic and socioeconomic characteristics. Thus, using zip code–level data can facilitate understanding of how the pandemic affected different communities and can help to identify vulnerable groups and tailor interventions to the unique needs of each community.

Table 3. Granularity of Data Sets

Data set	Time granularity	Level of observation
Mobility data set	Minute	Individual
POI data set	—	POI
Bank data set	Day	CBG
Survey data set	Day	Zip code
Case data set	Day	Zip code

Our utilization of a variety of unique data sets offers several notable contributions, which we outline here.

First, to the best of our knowledge, this is the first effort to combine proprietary data sources across both the public and private sectors in a timely, meaningful, and privacy-friendly manner. In this sense, it offers a novel pathway for data-driven policymaking. Second, we develop a robust method of harmonizing heterogeneous data sources at different levels of granularity for optimal policy analysis. Third, our data sets are updated with high frequency. The incorporation of such data provides a crucial real-time dimension to our analysis. This allows for the generation of timely policy recommendations, thus facilitating rapid responses to fast-changing situations. Therefore, we are able to provide relevant and timely advice to policymakers, potentially enabling more effective interventions during critical periods of a public health crisis.

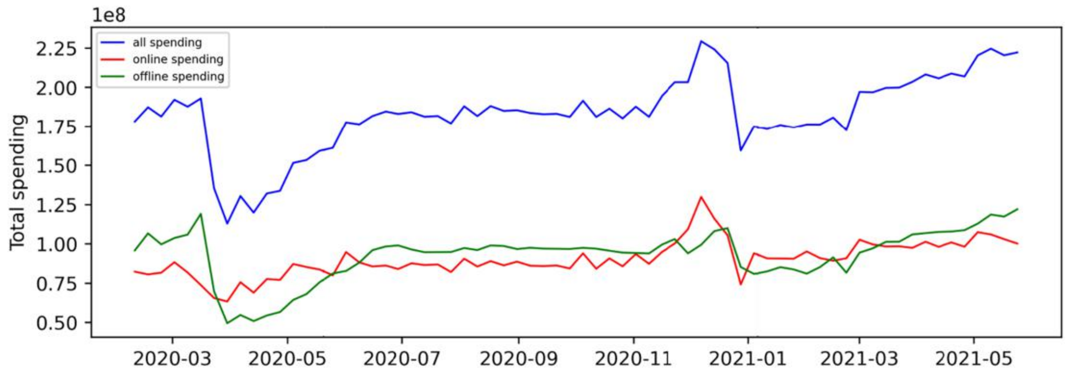
Together, these contributions reflect our commitment to integrating various data sets to deliver nuanced analysis and timely advice.

3.3. Model-Free Analysis

In the bank data set, COVID-19 has shown a great impact on the consumption structure of different income groups in Pennsylvania.

As shown in Figure 3, the weekly total spending dropped by almost half in late March and early April

Figure 3. (Color online) Spending Trend in Online and Offline Transaction Means from February 2020 to June 2021 in Pennsylvania



2020, when COVID-19 began to transmit massively within the United States. In the meantime, the online spending percentage of the total spending increased from 45% to more than 55%.

In some industry categories, the story confirms the previous trend. Take grocery stores as an example. In Figure 4, people in zip codes that have median income in the top 25% bracket are recognized as the "top income" group. People in zip codes that have median income in the bottom 25% bracket are recognized as the "bottom income" group. In both groups, the solid line represents offline spending, and the dashed counterpart represents online spending.

In Figure 4, we see a doubling in online grocery spending percentage after April 2020, when COVID-19 hit the United States. This phenomenon of spending migration from offline to online will be discussed in detail in Section 5. Figure 4 also illustrates the relative spending (Spending/Spending in Feb. 2020) trend for groceries. Although the relative spending with offline means maintains a stable level, the online relative spending almost doubles in the bottom income group and triples in the top income group. This not only echoes the online migration above but also raises the question of why it has a larger effect on the top-income group than the bottom-income group. This question will be explored in detail in Section 5.

However, in some industry categories, such as video games, the trend is different. In Figure 5, the spending on video games increased at the beginning of the pandemic (April 2020) and maintained itself at a stable level thereafter. This was possibly because video games are mostly purchased online, similar to the online spending trend (the bottom line in March 2020) in Figure 3. Another thing to note here is that the bottom income group has more increase in relative spending than the top income group during the pandemic, which is not seen in any other (70) industry categories in the bank data set.

A clearer view of this behavioral contrast between different income groups during the pandemic is provided by Figure 6. According to Risk Assessment Chart (Texas Medical Association 2020), we classified industry categories in bank data set into three

Figure 4. (Color online) Spending Trend in Grocery Store Category with Different Income Groups and Transaction Means from February 2020 to May 2021

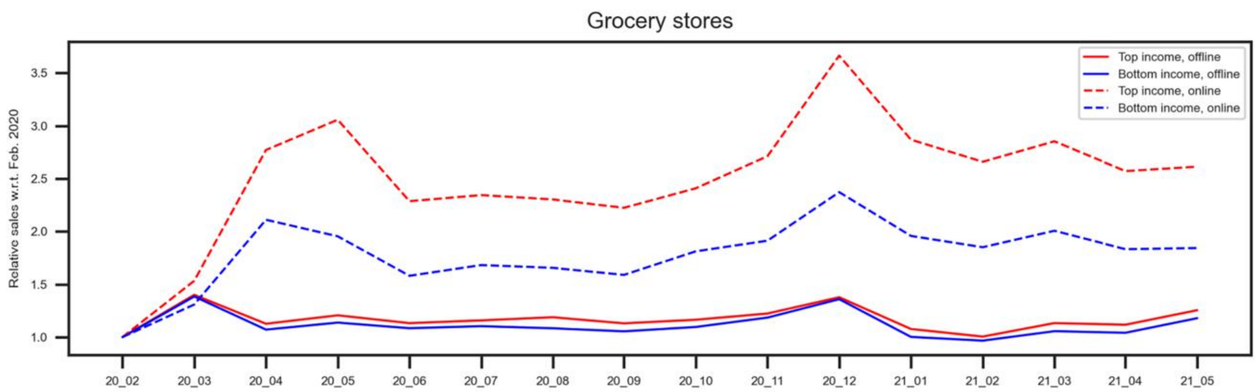
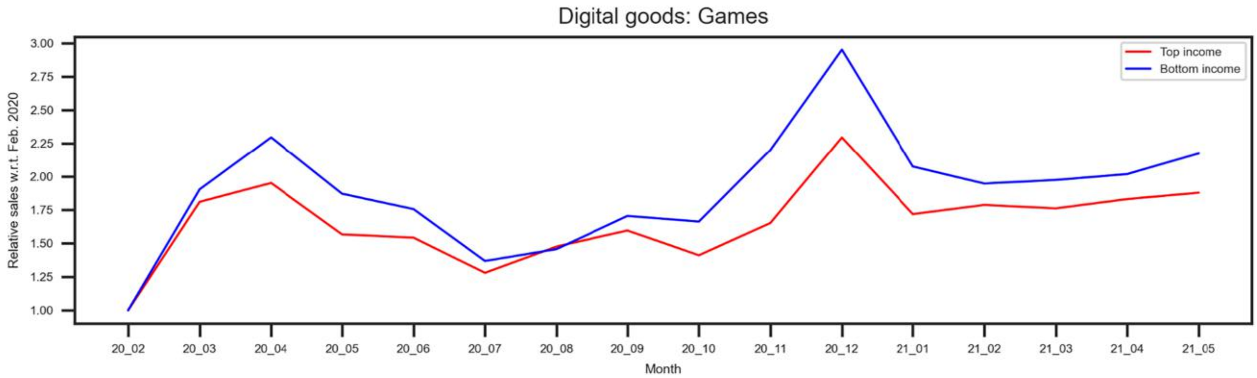


Figure 5. (Color online) Spending Trend in Video Game Category with Different Income Groups from February 2020 to May 2021



risk levels: high risk, moderate risk, and low risk. Solid lines in the top subfigure of 6 show the relative spending (Spending/Spending in Feb. 2020) of bottom income group, and the dashed lines show the relative spending of top income group, on three risk levels, respectively.

Meanwhile, both income groups show a drastic decrease in both moderate- and high-risk spending and a steady increase in low-risk spending; the degree of change is different.

The bottom subfigure of Figure 6 shows the ratio of relative change in spending (Relative spending of top income group/Relative spending of bottom income group). A higher ratio means spending change happens more in the top income group than in the bottom income group. It was found that the top income group

Figure 6. (Color online) Spending in Different Risk Levels, Conditioning on Income

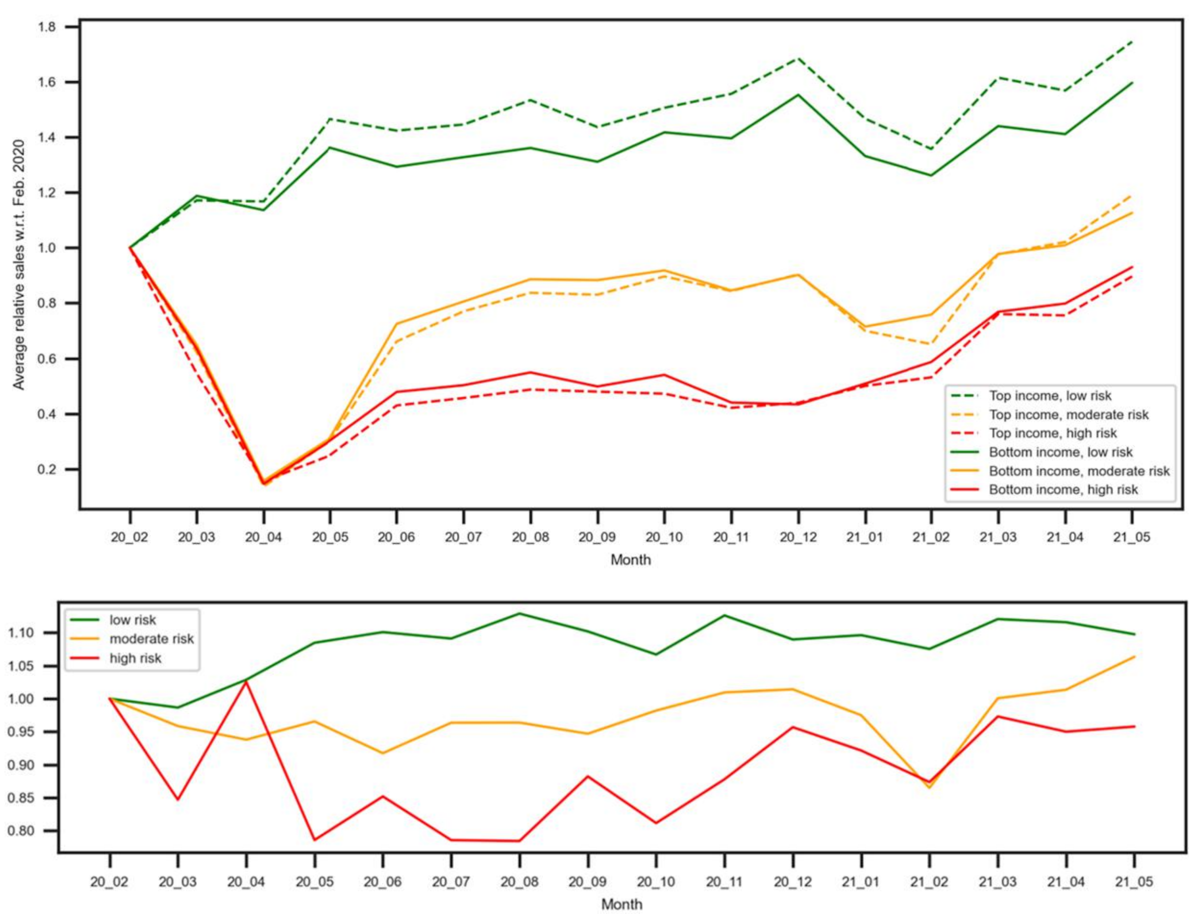
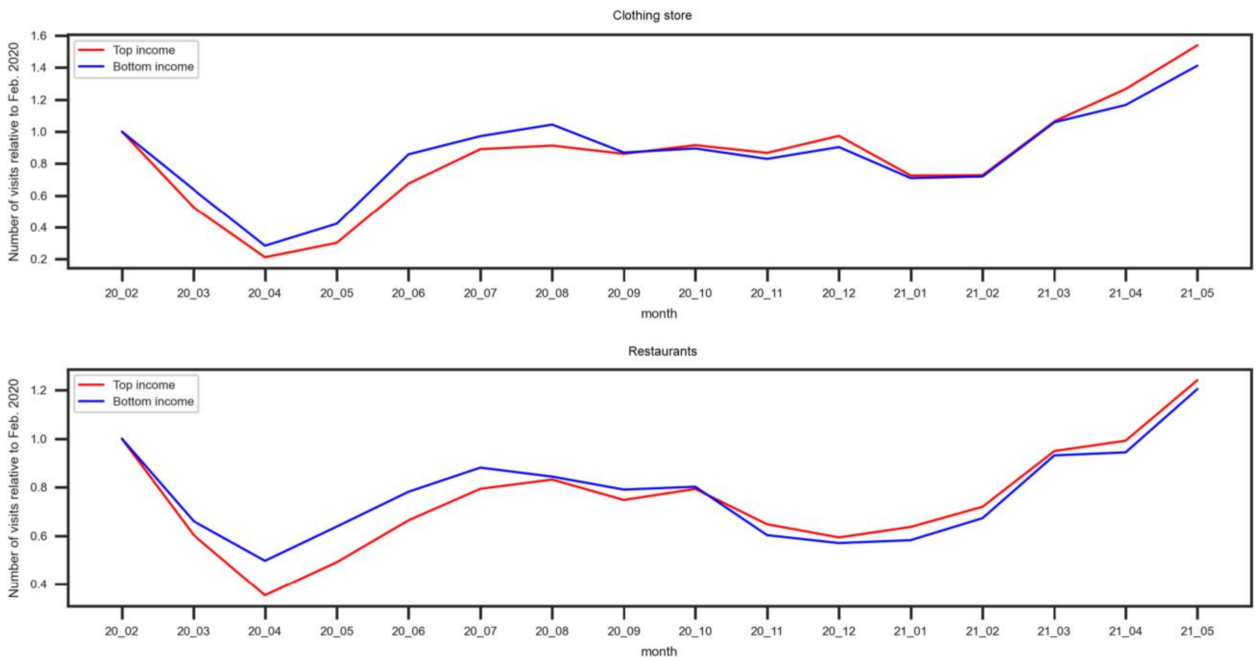


Figure 7. (Color online) Spending in Different Risk Levels, Conditioning on Income



increased their spending more than the bottom income group in low-risk activity whereas decreasing their spending more than the bottom income group in high-risk activity.

This phenomenon is also echoed by evidence from human mobility patterns revealed in *mobility data set*. In Figure 7, the high-income group (the bottom line in March 2020) relative to the low-income group (the top line in March 2020) exhibits a much more severe decrease in their visits to clothing stores and restaurants during the early months of COVID-19 and then a quicker recovery in mobility later in the year (i.e., after the reopening in the summer of 2020).

This evidence shows the behavioral differences between the two income groups. The bottom income group inevitably was exposed to greater health risk because of more offline spending on high-risk activities and more traveling overall, which raises the questions of what causes the difference and how the gap might be closed.

4. How Does Human Mobility Affect the Spread of COVID-19?

Existing epidemiological research mostly focuses on modeling the relationship between people's mobility pattern and their health risk from an *aggregate* level (e.g., country or state level). Scientists often have used aggregate-level mobility data to simulate the recent epidemic and estimate the parameters of the disease (Calafiore et al. 2020, Cooper et al. 2020).

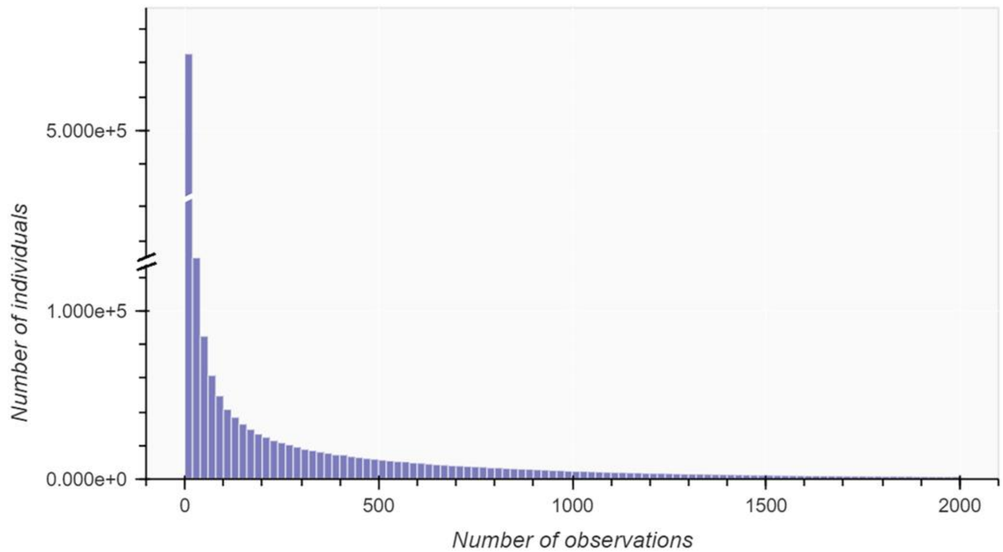
Modeling the spread of an epidemic from an aggregate level has a few drawbacks. First, although it aims to fit the overall data distribution from a population level, such a model cannot fully capture the data generation process from an individual perspective. Hence, it may lead to limited predictive power, especially given the strong heterogeneity in individual human behavior under different counterfactual scenarios. Second, mobility behaviors are, to a large extent, individual decisions. Models that focus solely on aggregated mobility trends present several drawbacks, limiting their effectiveness in understanding the underlying dynamics. Besides the limited interpretability, as they cannot explain what drives the observed changes, aggregated-level epidemiological models also suffer from a lack of precision and flexibility. Individual-level analysis, on the other hand, offers the opportunity to account for the heterogeneity in population behaviors, enabling more accurate predictions and targeted interventions. Furthermore, individual-level models can better capture the nuances in mobility patterns and their impact on disease transmission, thus providing valuable insights for policymakers and public health officials.

Facilitated by the unique and rich set of mobility data available at the individual level, we were able to develop a novel structural model of COVID-19 transmission that can help assess health risk (i.e., Susceptible, Infected, and Recovered probabilities) at the individual level.

4.1. Individual-Level SIR Model

The SIR model of disease spread is a three-stage differential equation model. It classifies people into three groups: susceptible, infected, and recovered. The susceptible group is defined as people who have not been

Figure 8. (Color online) Number of People with a Certain Number of Recorded Visits to POIs per Month in Mobility Data Set



infected, the infected group as those who have been infected and have not recovered, and the recovered group as those who have been infected but recovered.

The independent variable in the SIR model is time t, which is usually measured in days. On day t, the number of individuals in groups is $S(t)$, $I(t)$, and $R(t)$, respectively. Similarly, the fraction of individuals in groups is $s(t) = S(t)/N$, $i(t) = I(t)/N$, $r(t) = R(t)/N$, where N is the whole population. Note that $s(t) + i(t) + r(t) = 1$ for all t.

The main simplifications of the SIR model are that no immigration or birth is taken into account, no second-time infection is considered, no person is immune to the disease unless they are infected once, and the rate of change in $S(t)$, the number of susceptibles, depends only on $S(t)$, $I(t)$.

An important assumption the SIR model makes is that the rate of change in $S(t)$, $dS(t)/dt$, is dependent only on the interaction between susceptible and infected people and that in particular, each infected person has b contacts per day with other people. Because the infected people do not meet only susceptible people, the SIR model also assumes that the population is mixed homogeneously. Therefore, the susceptible equation is $dS(t)/dt = -bs(t)I(t)$.

However, the homogeneous mixing assumption is not true in reality. For example, large apartment buildings or multigenerational homes may have people in the same group living together. Also, visiting low-risk areas will incur exposure to fewer infected people than will visiting high-risk areas. Furthermore, a person may travel significantly more than other people because of various reasons, which can be evidenced in Figure 8 and Table 4. Socio-demographic differences are not taken into account in the original SIR model either.

To tackle the weakness of the SIR model, we propose the individual SIR model. We considered the spread of the disease, COVID-19 in this case, in a dynamic manner and took into account individual mobility and socio-demographic backgrounds.

At $t = 0$, each person $j \in J$, where J is the total population, gets assigned a probability of being in one of the three stages $p_{S_j}^{(0)}, p_{I_j}^{(0)}, p_{R_j}^{(0)}$. Note that $p_{S_j}^{(t)} + p_{I_j}^{(t)} + p_{R_j}^{(t)} = 1$ for all t.

A person is exposed to COVID-19 by visiting POIs and contacting infected people in those POIs. We denote the transition probability between two stages as $p_{S->I_j}^{(t)}$ and $p_{I->R_j}^{(t)}$. Although there are more transitions, for example, $R- > S$, according to the assumption the SIR model makes, the probabilities of transitions apart from $S- > I$ and $I- > R$ are zero.

Figure 9 shows this transition of stage, and the transition matrix for any individual is

$$\begin{pmatrix} 1 - P_{S->I_j}^{(t)} & 0 & 0 \\ P_{S->I_j}^{(t)} & 1 - P_{I->R_j}^{(t)} & 0 \\ 0 & P_{I->R_j}^{(t)} & 1 \end{pmatrix} \begin{pmatrix} p_{S_j}^{(t)} \\ p_{I_j}^{(t)} \\ p_{R_j}^{(t)} \end{pmatrix} = \begin{pmatrix} p_{S_j}^{(t+1)} \\ p_{I_j}^{(t+1)} \\ p_{R_j}^{(t+1)} \end{pmatrix}. \quad (1)$$

The probability of infection, $P_{S->I}$, is dependent on the exposure level of COVID-19, which can be impacted by various factors, including age, the number of people residing in the same household, the number of infected people visited the same POI within the same day, the type of the POI, and so on.

An individual j on the day t is observed to have visits to a (repeatable) list of POIs $(O_1, O_2, \ldots, O_n)$. Each POI in the list has an industry category preassigned to it by SIC code, denoted as c_i. For each POI c on day t, there is a set of people J_c visiting it. Then, for individual j, the risk of visiting POI O_1 on day t is proportion to the

Table 4. Inter–Zip Code Travel Frequency in Mobility Data Set

	Top income	Middle income	Bottom income
Top income	53.8%	22.7%	23.5%
Middle income	33.8%	30.5%	35.8%
Bottom income	28.1%	26.4%	44.4%

Notes. The number in row 1, column 2 is the proportion of travels from top-income zip codes to middle-income zip codes in all travels from top zip codes. Top-income zip codes are those in which the median income is in the top 25% of U.S. zip codes; bottom-income zip codes are those in which the median income is in the bottom 25% of U.S. zip codes, whereas middle-income zip codes are those in between.

expected number of infected people that also visit O_1 on day t.

The expected number of infected people that also visit O_1 on day t is

$$N_1^{(t)} = \sum_{j \in J_1} \rho_{O_1} p_{I_j}^{(t)}, \tag{2}$$

where ρ_{O_1} is the percentage of not wearing masks in the county that the POI O_1 belongs to.

The probability of infection, taking industry category and socio-demographic heterogeneity into account, is

$$p_1 = (\beta_2 Cat_1 + \beta_3 Age + \beta_4 LgInc)\rho_1 N_1^{(t)}, \tag{3}$$

where Cat is the dummy variable for the industry category of the POI, Age is the age of the individual, $LgInc$ is the (log) household income of the individual, and ρ_1 is the percentage of not wearing masks in the county that the individual j is from.

The individual also has an unobserved infection event at home, and the probability of infection is

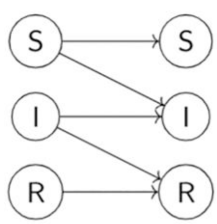
$$p_{unob} = \beta_1 N_{hh}, \tag{4}$$

where N_{hh} is the household size.

If we assume each event of potential infection is independent, the probability of infection on day t is

$$P_{S->I_j}^{(t)} = 1 - (1 - p_{unob}) \prod_{i=1}^{n} (1 - p_i) \tag{5}$$

Figure 9. Stage Transition in Individual SIR Model



$$= 1 - (1 - \beta_1 N_{hh}) \prod_{i=1}^{n} \left[1 - (\beta_{2Cat_1} + \beta_3 Age + \beta_4 Inc)\rho_1 \sum_{j \in J_i} \rho_{O_1} p_{I_j}^{(t)} \right]. \tag{6}$$

The transition probability from infected to recovered is

$$P_{I->R_j}^{(t)} = 1/\delta_I, \tag{7}$$

where δ_I is the expected days it takes to recover from COVID-19. As in Buckee et al. (2020), $\delta_I = 4$.

4.2. Empirical Analysis

The mobility data set is an anonymous data set and does not contain demographic information such as age or income. Therefore, we augmented the mobility data set with a synthetic population data set, namely the FRED (Framework for Reconstructing Epidemic Dynamics) synthetic population data set (Grefenstette et al. 2013).[2] The synthetic population data set had been created using an iterative approach to U.S. Census Data. The synthetic population data set contains a population with characteristics sampled from distributions of those in the census data, that is, each individual in the synthetic population data set has a home address, age, income, household size, and so on, assigned, and the aggregated statistics of the characteristics are close to the census data, visualized in Figure 10.

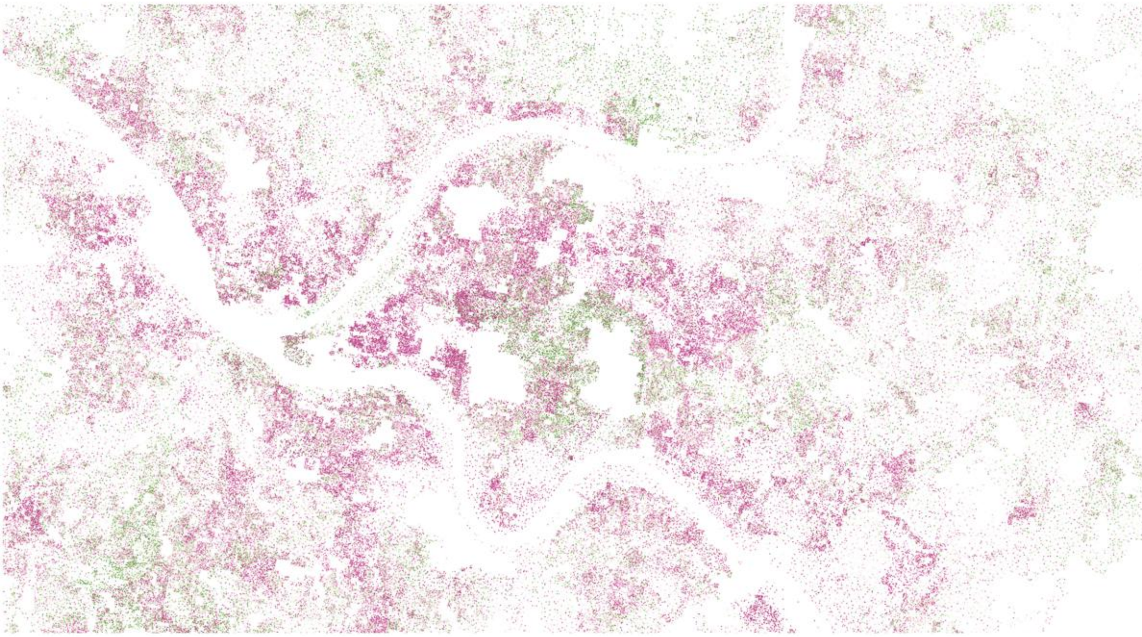
To map the synthetic population data set onto the mobility data set, we identified the most frequently visited place as the home address of an individual and uniformly sampled within 100 meters from the synthetic population data set. As the demographic distribution of the mobility data set is unknown, we validated it with several different nonuniform distributions in Appendix A and thereby confirm the robustness of this method.

To estimate the individual SIR model, we simulated the spread of COVID-19, predicted the case numbers of each county in Pennsylvania, and calculated the root mean square error (RMSE) of the predicted case number with respect to the ground truth number in the *case data set* as an objective function:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(N_{\text{cases}}^{(t)} - \hat{N}_{\text{cases}}^{(t)} \right)^2}. \tag{8}$$

We deployed a stochastic gradient descent (SGD) method to estimate the model. To be specific, we used the adaptive moment estimation (Adam) optimizer to train the model with an initial learning rate of 1 e-3. We also adopted the early-stopping technique to avoid overfitting. The parameter estimations for the individual SIR model are in Table 5.

Figure 10. (Color online) Visualization of FRED Synthetic Population Data of Allegheny County, Pennsylvania



Notes. Each dot represents an individual on the map. Deeper color represents higher population density.

The model was validated by predicting the case number in Pennsylvania in both an out-of-sample fit and a full fit. We also compared our model with the classic SIR model without mobility data, as shown in Figure 11. We used the whole data set as training data and tested it on the same data set in full fit, whereas in the out-of-sample fit, we trained the model only with data before October 11, 2020 and then tested it on the whole data set until October 29, 2020. Therefore, the out-of-sample fit was an evaluation of the ability of the model using mobility data to predict the case number accurately in an unseen time period. The experiment documented a 40% drop in RMSE in out-of-sample tests after integrating individual mobility information.

The out-of-sample test showed that our model fits the pandemic case well. The negative sign of β_4 shows that the risk of exposure is negatively correlated with household income and that the positive sign of β_1 and β_3 shows that the risk of exposure is positively correlated to the size of household and age, holding all other variables constant.

Meanwhile, Figure 12 shows the estimated category risk compared with the Risk Assessment Chart (Texas Medical Association 2020). The Kendall Tau distance between the risk ranking of our estimation and that of the Risk Assessment Chart was $3/21 = 0.14$, indicative of high consistency.

Table 5. Parameter Estimation of Individual SIR Model

Parameters	Estimate
β_1 (size of household)	0.0152
β_3 (age)	0.0075
β_4 (log monthly income)	−0.102

5. How Does COVID-19 Risk Affect Shopping Behavior?

We considered a binary choice model in which the consumer chooses to purchase an item of category c via online or offline transaction means by considering the potential risk of exposure at the target offline merchant r if choosing to shop offline, the delivery fee f if choosing to shop online, the travel distance from home to the merchant d, whether the day is a weekend w, and the income s of the consumer.

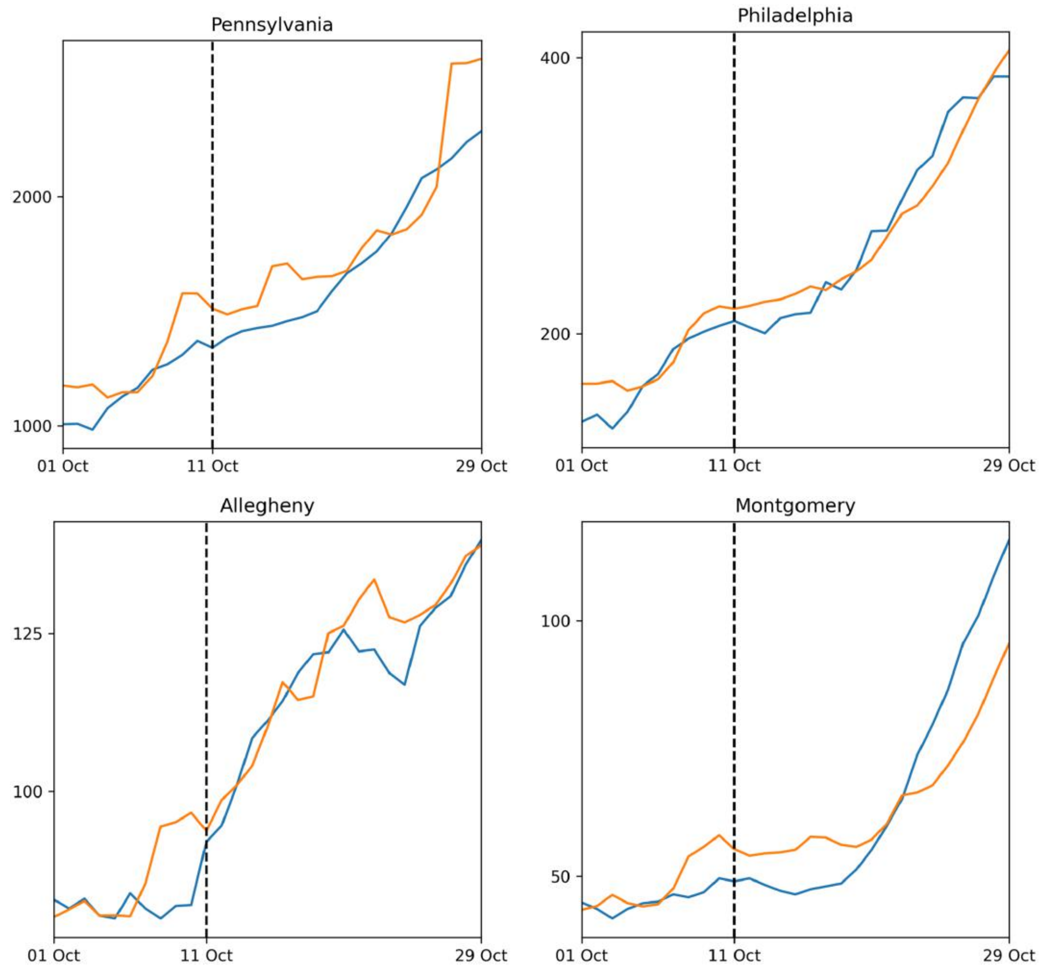
The potential infection risk of an offline purchase is dependent on three factors, as discussed here.

An inherently more dangerous industry category can render the consumer more exposed to COVID-19. For example, dining at a restaurant is much more risky than filling a car up at a gas station. The risk factor of the industry category is estimated via the individual SIR model and is denoted as ρ_c, where c is the industry category.

The expected number of people visiting a location is also an important factor in evaluating risk. To estimate it, we used the daily average number of people who made offline purchases at merchants of the same category c in the same zip code i in the previous week:

$$\hat{N}_{c,i}^{(t)} = \frac{1}{7}\sum_{d=t-7}^{t-1} N_{c,i}^{(d)}. \tag{9}$$

Figure 11. (Color online) COVID-19 Case Number Out-of-Sample Prediction Test



Notes. Two lines represent the predicted and reported case number. We fit the model with mobility data from October 1 to October 11, 2020, and predict the case numbers from October 1 to October 29, 2020, in the whole of Pennsylvania and its three most populous counties: Philadelphia, Allegheny, and Montgomery.

The last factor is the infection rate of the zip code. The infection rate is defined by

$$\delta_i^{(t)} = \frac{N_{I,i}^{(t)}}{N_i^{(t)}}, \tag{10}$$

where $N_{I,i}^{(t)}$ is the number of reported infected people in zip code i, and $N_i^{(t)}$ is the total number of people in zip code i.

Hence, the potential infection risk of an offline purchase at merchant of category c and zip code i and day t is

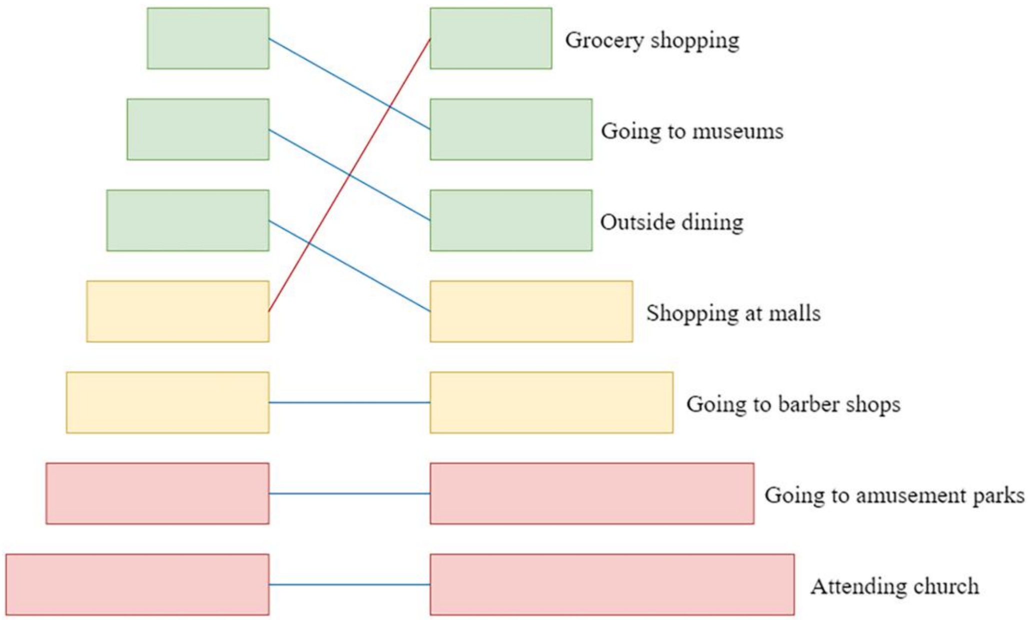
$$r_{c,i}^{(t)} = \rho_1 \times \rho_2 \times \rho_c \times \hat{N}_{c,i}^{(t)} \times \delta_i^{(t)}. \tag{11}$$

In Equation (11), $\hat{N}_{c,i}^{(t)}\delta_i^{(t)}$ is the estimated number of infected people, and ρ_c is a modifier for industry category of the merchant, which is consistent with Equation (6) in Section 4; ρ_1 is the percentage of not wearing masks in the county that the offline merchant belongs to, whereas ρ_2 is the percentage of not wearing masks in the county of the customer. Both variables are from the survey data set.

The delivery fee is the main reason consumers hesitate to shop online (Huang and Oppewal 2006, Bauerová 2018, Dias et al. 2021); therefore, we integrated "delivery fee" as a variable in the behavior model. We assumed that the delivery fee is proportional to the product of the transaction amount and the expected travel distance. Although many of the grocery and restaurant delivery platforms offer a monthly subscription and flat delivery fee promotions, the service fee and default tips are usually proportional to the total transaction amount and the expected travel distance.

The exact travel distance was unobserved because the consumer j's home address or the address of the merchant was not directly observed. Therefore, we made the assumption that the merchants of the same industry category c and zip code i are distributed

Figure 12. (Color online) Categorical Risk Validation



Notes. Left bars are the estimated category risk; right bars are the category risk according to the Texas Medical Association (2020) Risk Assessment Chart. The line shows the correct ranking pair, whereas the red line shows the wrong ranking pair. The Kendall Tau distance between two arrangements of risk levels is $3/21 = 0.14$.

randomly in a square. This is a data limitation that could be potentially addressed in future studies through more precise data collection methods or advanced estimation techniques, where the home address and merchant address are both known in a transaction record.

The consumer's home address is also randomly distributed in the same square. We calculate the expected distance to the closest merchant as d:

$$d_i^{(t)} \approx \frac{\sqrt{S_i}}{\sqrt{N_{c,i}}}; \tag{12}$$

S_i is the area of the zip code i, and $N_{c,i}$ is the number of merchant of category c in zip code i. An important assumption we made is that the consumer will visit the merchant in the same zip code. Thus, we can use the median household income of the zip code on the transaction record as a proxy variable for the income of the consumer. One argument for this approximation is that although consumers will inevitably travel to other zip codes, they will travel to the zip code that has a similar income level more often, as shown in Table 4. Therefore, using the income of the merchant's zip code is a reasonable proxy for the income of the consumer's income.

The linear probability model is

$$P_{offline} = P(y = 1|x) = \beta_{const} + \beta_{risk} r + \beta_{deliv} f + \beta_{dist} d + \beta_{amt} a + \beta_{inc} s, \tag{13}$$

where $y = 1$ if and only if the consumer chooses to purchase offline.

The estimation of the model is on bank data set, using data in the top eight industry categories (in total transaction amount) from April 2020 to October 2020. Industry categories that are offline only are ignored, for example, gas stations. We report the estimation result in Table 6.

6. Equilibria and Counterfactual Analyses

To reveal the effect of COVID-19 policies on the overall risk of COVID-19 and the economy by conducting counterfactual simulations, we first establish equilibria from the mobility model and the behavior model.

The equilibrium underlies the choice process of online purchase. When choosing to make a purchase of the same good via online or offline transaction means, the consumer would consider several factors discussed in Section 5. The negative feedback loop is as follows:

Table 6. Parameter Estimation of Choice Model

Parameters	Estimate	Std. error
β_{risk}	-6.13×10^{-3}***	1.46×10^{-3}
β_{deliv}	9.64×10^{-4}***	1.86×10^{-4}
β_{dist}	-2.83×10^{-2}***	2.04×10^{-3}
β_{amt}	7.37×10^{-3}***	2.85×10^{-3}
β_{inc}	-2.36×10^{-6}***	1.46×10^{-8}

*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$.

Consumers are more willing to purchase offline (for some reason) → there will be more people in the offline places → the risk of exposure to COVID-19 goes up → consumers are less willing to purchase offline.

The equilibrium is characterized by the optimization equation:

$$\min_{p_{offline}} \quad \delta_{infect}(p_{offline}; r, f, d, w, s, c)$$
$$s.t. \quad p_{offline} > 0, p_{offline} < 1$$
$$p_{offline} = g(\delta_{infect}).$$

In the equilibrium, we use the infection rate of a zip code δ_{infect} and the probability of consumers going offline $p_{offline}$ as the two economic variables.

In the individual SIR model, the relationship between δ_{infect} and $p_{offline}$ is controlled by the expected number of infected people at POIs $\sum_{j\in J_i} p_{I_j}^{(t)}$ in Equation (6):

$$\delta_{infect}^{(t+1)} = (1-\delta_{infect}^{(t)})P_{S->I_j}^{(t)} - \delta_{infect}^{(t)} P_{I->R_j}^{(t)} \tag{14}$$

$$= C_1 P_{S->I_j}^{(t)} + C_2 \tag{15}$$

$$= C_1\left[1-(1-\beta_1 N_{hh})\prod_{i=1}^{n}\left[1-(\beta_{2Cat_1}+\beta_3 Age + \beta_4 Inc)\sum_{j\in J_i} p_{I_j}^{(t)}\right]\right] + C_2 \tag{16}$$

$$\approx C_1\left[\beta_1 N_h h + \sum_{i=1}^{n}(\beta_{2Cat_1}+\beta_3 Age + \beta_4 Inc)\right]\sum_{j\in J_i} p_{I_j}^{(t)} + C_2 \tag{17}$$

$$= C_1^* p_{offline}^{(t)} + C_2^*. \tag{18}$$

This equation has shown the quasi-linear relation between δ_{infect} and $p_{offline}$. Equation (16), where C_1 and C_2 are constants, holds because we treat $\delta_{infect}^{(t)}$ as a constant (we cannot change the past). The approximation in 18 is made for the fact that every single event of infection only has a probability that is significantly smaller than one. The relation between δ_{infect} and $p_{offline}$, the infection curve, is illustrated in Figure 13(a).

Conversely, in the consumer behavior model discussed in Section 5, the customer will have less chance to make purchases offline in fear of infection.

In Equation (13), we have

$$p_{offline} = \beta_{const} + \beta_{risk} r + \beta_{deliv} f + \beta_{dist} d + \beta_{weeknd} w + \beta_{inc} s + \sum_{c\in C\backslash c_0} \beta_c c \tag{19}$$

$$= C_2 + \beta_{risk} r \tag{20}$$

$$= C_2 + \beta_{risk}\rho \hat{N}\delta_{infect} \tag{21}$$

$$= C_1\delta_{infect} + C_2. \tag{22}$$

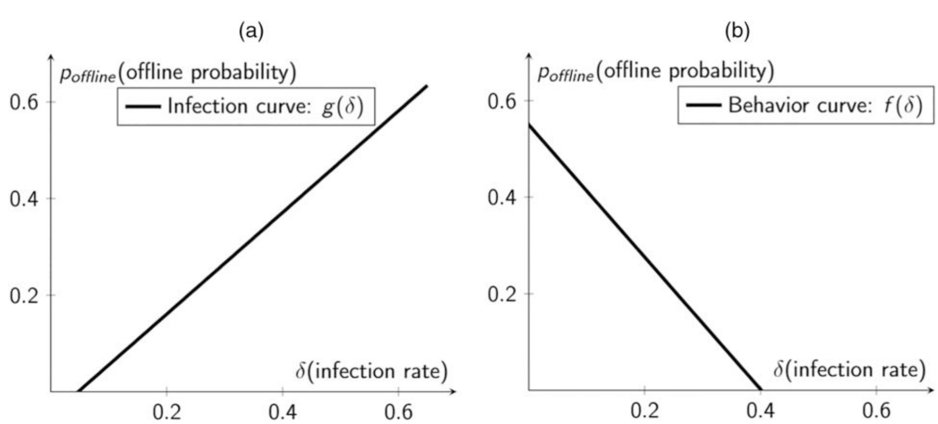
Because $\beta_{risk} < 0$, $\hat{N} > 0$, $\rho > 0$, $C_2 < 0$. Figure 13(b) shows the behavior curve.

The intersection of the infection curve and the behavior curve is the equilibrium point and is also where the reality is achieved at day t. Policymakers can change variables in the feedback loop and shift the infection or behavior curve to achieve a better balance between the economy and public health risk.

6.1. Would Delivery Fee Coupons Help the Poor?

The delivery fee can be a major factor preventing people from using the online purchasing channel (Dias et al. 2021). By distributing delivery coupons, the behavior curve will shift left because people will be more

Figure 13. Infection and Behavior Curve



Notes. (a) The infection curve. (b) The behavior curve. Both shown as linear functions for simplicity.

willing to purchase online under the same infection rate. As indicated in Figure 14, the left shift of the behavior curve will cause the equilibrium point to shift left and down, which means that both the probability of making an offline purchase and the infection rate go down at the same time.

Furthermore, although we did not observe people from low-income zip codes visit POIs of inherently more dangerous categories, we did observe they have a more limited choice of POIs to visit. Thus, the POIs become more crowded, which is reflected in $\sum_{j\in J_i} p_{I_j}^{(t)}$. Having a greater $\sum_{j\in J_i} p_{I_j}^{(t)}$ means having a more gentle slope of the infection curve, as shown (exaggeratedly) in Figure 15. Even if delivery coupons are given out in the same amount, the poor will benefit more from it.

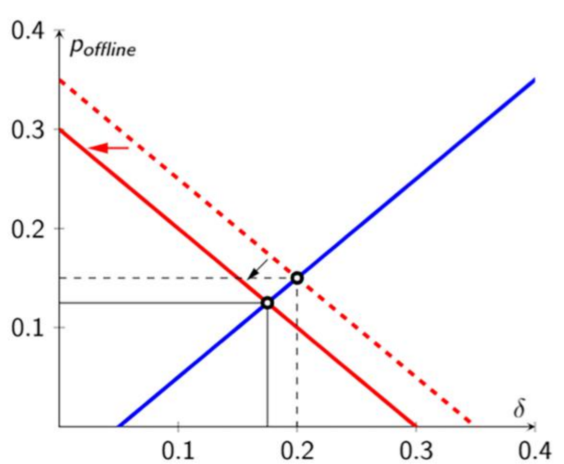
In our simulation, giving out delivery coupons for every one dollar's good (consumers do not have to pay for the delivery fee for one dollar's good) will decrease (immediate) infection rate of the general population by 7×10^{-6}, which accounts for 90 people per day in April 2020 in Pennsylvania. Meanwhile, giving out delivery coupons for every one dollar's good to people living in the lower 25% income zip code will decrease (immediate) infection rate by 9×10^{-6}, which accounts for 115 people in Pennsylvania, as in Table 7.

The long-term effect of a decreased infection rate is compounded over time; giving out delivery coupons can be much more beneficial than we estimated. We suggest that the government distribute delivery coupons, especially to the lower-income districts.

6.2. Is the Mask Mandate Effective?

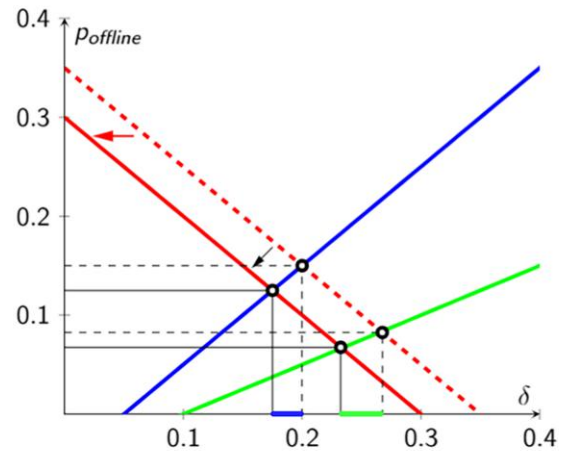
Mask mandate is a policy of significance during the pandemic. By enforcing the mask mandate, the industry-category risk will go down. Thus, the infection will be less likely to happen with the same number of visits, and people will worry less about the potential danger of exposure when shopping offline.

Figure 14. (Color online) Delivery Fee Coupon Helps



Note. Dashed line: behavior curve before the intervention; solid line with negative slope: behavior curve after the intervention; solid line with positive slope: infection curve.

Figure 15. (Color online) Delivery Fee Coupon Helps Differently



Notes. Dashed line: behavior curve before intervention; solid line with negative slope: behavior curve after intervention; top line with positive slope: infection curve of high-income group; bottom line with positive slope: infection curve of low-income group. The change in the infection rate of the low-income group is greater than that of the high-income group.

As in Figure 16, the infection curve will rotate counterclockwise, and the behavior curve will rotate counterclockwise as well. In the experiment, we applied a flat discount to each industry-category risk as the proxy of the mask mandate, and we found that, although the mask mandate can reduce the infections per visit (by definition), the incentive of the mask mandate for people to do more in-store purchasing can offset this effect. To be specific, without any changes in other variables, if the percentage of those not wearing a mask is reduced by half, there will be an estimated 5.3% drop in COVID-19 infections in Pennsylvania but a 12.4% increase in offline shopping as well.

However, as we will point out, this is not always the case.

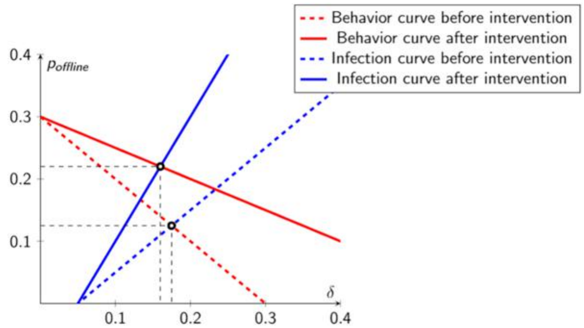
Figure 17 illustrates this phenomenon. Initially, the absence of a mask mandate policy is depicted by the dotted curves, representing the behavior and infection rates, respectively. In this scenario, people estimate a higher risk and therefore prefer online shopping, minimizing the potential for infections.

As Intervention I is implemented, denoted by the dashed curves, a stricter mask mandate policy leads to a marked-down estimated risk. This effect influences

Table 7. Estimated Infection Rate Decrease for Different Income Groups with Delivery Coupon

Overall	Top 25% income	Bottom 25% income
7.03×10^{-6}	5.28×10^{-6}	9.45×10^{-6}

Figure 16. (Color online) Movement of Equilibrium After the Intervention

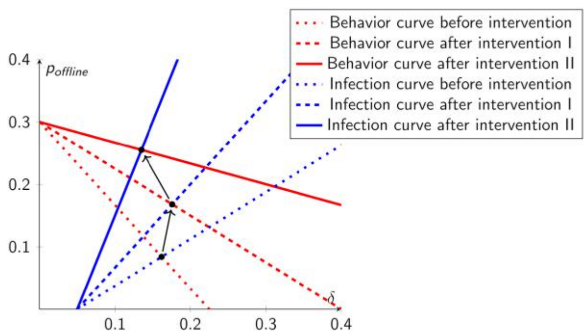


individuals to shop offline more frequently. Interestingly, this increase in offline shopping does not correspond to a decrease in infections. Rather, the opposite happens. Because of the increased public exposure, even with a stricter mask mandate, there is an unexpected rise in infections. This is a manifestation of the "risk compensation" effect, whereby perceived safety measures can lead to riskier behavior.

As illustrated by the solid curves, Intervention II represents a further tightening of the mask mandate. In this case, the protection provided by the mask mandate outpaces the number of people venturing into public spaces, resulting in a decline in infection rates. People still feel safe enough to continue their offline shopping, but the strengthened mask mandate effectively curbs the virus spread, proving that a balance can be achieved between safety measures and maintaining a degree of normalcy in daily activities.

Although in the real-life setting, we observed infection rate reductions only under a tighter mask mandate policy, in essence, these hypothetical interventions elucidate the nuanced relationship between public health policy, individual behavior, and infection rates. The effectiveness of policies may not always align with initial expectations because of adaptive changes in human behavior, which underlines the importance of ongoing evaluation and adjustment of policies as they are implemented.

Figure 17. (Color online) Equilibrium Exhibits a Right-Then-Left Shift as Depicted in the Graph



6.3. Inequality in Categorical Lockdown Effect

Lockdown of POIs of certain categories has been a widely used policy in the past two years (Abu-Rayash and Dincer 2020, Badr et al. 2020, Kraemer et al. 2020, Chang et al. 2021). The main aim of the lockdown, especially in the high-risk category, for example, restaurants, movie theaters, and so on, is to block the critical path in the disease spread trajectory. The effectiveness of this treatment has been analyzed in Le et al. (2022) and Chang et al. (2021). However, as discussed previously, the lack of individual-level mobility data and, more importantly, the lack of any substitute for the loss of offline activity are the main shortcomings of such studies.

In an attempt to address these issues, we conducted two counterfactual tests. First, we simulated the lockdown effect solely with the individual SIR model. Using mobility data set during February 2020, we block certain categories of POIs in the travel trajectories and calculate the risk of infection, with the coefficients estimated in the original model, ignoring the blocked waypoints as shown in Equation (24):

$$P_{S->I_j}^{(t)} = 1 - (1 - p_{unob})\prod_{i=1}^{n}(1 - p_i) \tag{23}$$

$$= 1 - (1 - \beta_1 N_{hh})\prod_{i=1}^{n}\left[1 - (\beta_{2Cat_1} + \beta_3 Age + \beta_4 Inc)\rho_1 \sum_{j\in J_i} \rho_{O_1 \backslash O_{ignore}} p_{I_j}^{(t)}\right], \tag{24}$$

where O_{ignore} is the set of the ignored waypoints in the trajectories.

We tested a complete lockdown scenario on three different categories: restaurants, department stores/shopping malls, and grocery stores. We found that the total infection rate decreased in Pennsylvania. With the cutoff of the crucial disease-spreading path, the infection rate became evener across all zip codes in Pennsylvania.

In a more elaborate attempt, we took into account customers' behavior in opting for online shopping after the offline alternative was forbidden by the lockdown policy. We used the coefficients estimated in the previous section with the *mobility data set* and *bank data set*

Figure 18. (Color online) Lockdown Intervention

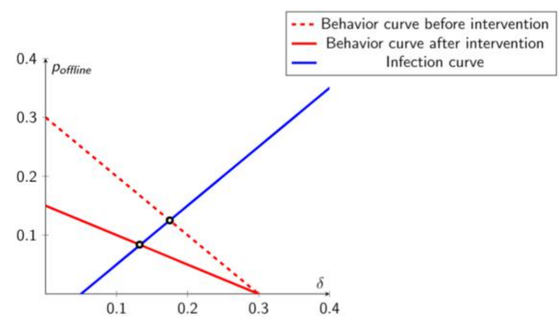


Table 8. Estimated Infection Rate Decrease for Different Categories to Lockdown

	Estimated infection rate decrease		
Category to lockdown	Overall	Top 25% income	Bottom 25% income
Restaurants	13.5%	17.0%	11.0%
Grocery stores	7.9%	9.1%	6.0%
Department stores	5.3%	5.9%	4.2%

Note. The top income group saw a more significant effect on lockdowns.

from February 2020. The experiment revealed a smaller total infection rate reduction than in the pure epidemiological setting. The lockdown policy was less effective, mainly because of customers having not completely avoided the blocked categories, as shown in Figure 18. We estimated a 33.8% infection rate reduction for the top 25% income zip code and a 25.8% drop for the bottom 25% income zip code in the event that the top three categories (grocery stores, restaurants, and department stores) were locked down, as in Table 8.

7. Discussion and Conclusion

We discussed and proposed herein a holistic model for joint modeling of human mobility, health risk, and economic activity on two unique large-scale data sets. We broke down the model into three parts.

First, we introduced a fine-grained individual susceptible-infected-recovered model (Individual SIR Model) to a large GPS location tracking data set to estimate the impact of human mobility on the spread of COVID-19. We found that household size and an individual's age had a positive correlation with COVID-19 spread, whereas income had a negative correlation. Second, we proposed a binary choice model on a large credit card data set to capture the mobility response to COVID-19. We found that delivery fee and weekend effects had a significantly negative impact on online purchase probability, whereas exposure risk of COVID-19, average distance from POI, and income level had a significantly positive impact on online purchase probability. Last, we set up an equilibrium based on the negative feedback loop of the mutual effects of COVID-19 spread and human activity. We tested three counterfactual policies: giving out delivery coupons, enforcing mask mandates, and announcing categorical lockdowns. We found that giving out delivery coupons would reduce the infection rate and have greater effects on low-income groups. Interestingly, we also found that whereas mask mandates, in theory, will curb virus spread, loose enforcement may unintentionally increase the overall infection rate. In the categorical lockdown setting, we found that a lockdown in Pennsylvania, encompassing restaurants, shopping malls, and grocery stores, led to a uniform reduction in infection rates across all zip codes, but its effectiveness was slightly diminished when taking into account a shift to online shopping.

Overall, our approach proved distinctive in its capacity to assemble, at the zip code level, individual-level data sets that had not been used together. These data enabled customization of the epidemic models widely used in public health with individual-level data traces of mobility behaviors for assessment of public health risks, which in turn enabled parameterization of economic choice models of how individuals make economic decisions. Various policy interventions and their capacities to shift the equilibrium between economic activity and public health were investigated in this study. Whereas the data-informed joint modeling approach was developed and tested in the pandemic context, it is generalizable for the evaluation of any counterfactual policy interventions.

Our work has some limitations that nonetheless can illuminate potentially fruitful paths for future research. First, we focused primarily on consumer behavior and did not incorporate any analysis of worker behavior. Although our research provides valuable insights into consumer preferences and decision-making, understanding the role of workers in the system could offer additional perspectives and potential optimizations. In future work, it would be beneficial to explore the influence of worker behavior, as it can provide a more comprehensive understanding of the ecosystem and inform strategies for improving overall efficiency. Additionally, we could extend our framework by developing a more detailed online delivery policy to address specific scenarios and optimize coupon utilization. Furthermore, expanding the merchant categorization within our framework would enable more nuanced analysis and, therefore, more tailored policies for lockdowns.

Appendix A. Robustness Check

Different distributions of possible socioeconomic populations are tested.

Table A.1 shows the estimated β_4 under different settings of normal distribution.

Table A.2 shows the relative deviation of $\beta_4 = \frac{\beta_4 - \overline{\beta}_4}{\overline{\beta}_4}$:

Beta distribution: $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$.

Six different Beta distribution settings are tested for the marginal distribution of age and income, as in Table A.3. It shows acceptable robustness under all settings.

For all four parameters, the "worst" distribution is when both marginal distributions have $\alpha = 0.5$ and $\beta = 0.5$, as seen in Table A.4. The reason might be when $\alpha = 0.5$ and $\beta = 0.5$,

Table A.1. Estimation of β_4 (Weight of Income)

Age mean/income mean	0.25	0.5	0.75
20	−0.129	−0.134	−0.107
40	−0.145	−0.121	−0.136
60	−0.118	−0.128	−0.095

Table A.2. Relative Deviation of β_4 (Weight of Income)

Age mean/income mean	0.25	0.5	0.75
20	0.075	0.117	0.108
40	0.208	0.001	0.133
60	0.002	0.007	0.208

Table A.3. Beta Distribution Settings Tested (Marginal)

α	β
0.5	0.5
5	1
5	2
2	2
2	5
1	5

Table A.4. Maximum Relative Deviation and Respective Joint Distribution

Parameters	Maximum relative deviation	α_{age}	β_{age}	α_{income}	β_{income}
β_1	0.215	0.5	0.5	0.5	0.5
β_2	0.321	0.5	0.5	0.5	0.5
β_3	0.309	0.5	0.5	0.5	0.5
β_4	0.364	0.5	0.5	0.5	0.5

the distribution assumes that we should sample the extreme value more (i.e., mostly very old, very young, very rich, and very poor people are presented in the *mobility data set*), which is not very probable.

Endnotes

[1] The mobility data set is an anonymous data set that does not contain demographic information such as the age and income of individuals. To analyze the heterogeneous effects on different demographic groups in terms of human mobility and health risk, we augmented the mobility data set with a synthetic population data set. We used the standard FRED synthetic population data set (Grefenstette et al. 2013). We discuss more details in Section 4.2.

[2] As a robustness check, we created a synthetic population data set in PA with a similar approach in Xu et al. (2017) and got similar results.

References

Abu-Rayash A, Dincer I (2020) Analysis of mobility trends during the COVID-19 coronavirus pandemic: Exploring the impacts on global aviation and travel in selected cities. *Energy Res. Soc. Sci.* 68:101693.

Adams-Prassl A, Boneva T, Golin M, Rauh C (2020) Inequality in the impact of the coronavirus shock: Evidence from real time surveys. *J. Public Econom.* 189:104245.

Alon T, Doepke M, Olmstead-Rumsey J, Tertilt M (2020) The impact of COVID-19 on gender equality. NBER Working Paper No. 26947, National Bureau of Economic Research, Cambridge, MA.

Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM (2020) Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *Lancet Infectious Diseases* 20(11):1247–1254.

Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A (2010) Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *J. Comput. Sci.* 1(3):132–145.

Bauerová R (2018) Consumers' decision making in online grocery shopping: The impact of services offered and delivery conditions. *Acta University Agric. Silvic. Mendel. Brun.* 66(5):1239–1247.

Bell DR, Song S (2007) Neighborhood effects and trial on the Internet: Evidence from online grocery retailing. *Quant. Marketing Econom.* 5:361–400.

Benzell SG, Collis A, Nicolaides C (2020) Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of us locations. *Proc. Natl. Acad. Sci. USA* 117(26):14642–14644.

Birge JR, Candogan O, Feng Y (2022) Controlling epidemic spread: Reducing economic losses with targeted closures. *Management Sci.* 68(5):3175–3195.

Block P, Hoffman M, Raabe IJ, Dowd JB, Rahal C, Kashyap R, Mills MC (2020) Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behav.* 4(6):588–596.

Bonaccorsi G, Pierri F, Cinelli M, Flori A, Galeazzi A, Porcelli F, Schmidt AL, et al. (2020) Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl. Acad. Sci. USA* 117(27):15530–15535.

Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, Grad YH, et al. (2020) Aggregated mobility data could help fight COVID-19. *Science* 368(6487):145–146.

Calafiore GC, Novara C, Possieri C (2020) A modified sir model for the COVID-19 contagion in Italy. *2020 59th IEEE Conf. Decision Control* (IEEE, Piscataway, NJ), 3889–3894.

Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J (2021) Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589(7840):82–87.

Chen Y-C, Lu P-E, Chang C-S, Liu T-H (2020) A time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Trans. Network Sci. Engrg.* 7(4):3279–3294.

Chowkwanyun M, Reed AL Jr (2020) Racial health disparities and COVID-19-caution and context. *New England J. Medicine* 383(3): 201–203.

Coibion O, Gorodnichenko Y, Weber M (2020) Labor markets during the COVID-19 crisis: A preliminary view. NBER Working Paper No. 27017, National Bureau of Economic Research, Cambridge, MA.

Cooper I, Mondal A, Antonopoulos CG (2020) A sir model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* 139:110057.

Courtemanche CJ, Le AH, Yelowitz A, Zimmer R (2021) School reopenings, mobility, and COVID-19 spread: Evidence from Texas. NBER Working Paper No. 28753, National Bureau of Economic Research, Cambridge, MA.

de Palma A, Vosough S, Liao F (2022) An overview of effects of COVID-19 on mobility and lifestyle: 18 months since the outbreak. *Transportation Res. Part A Policy Practice* 159:372–397.

Dias EG, Oliveira LKD, Isler CA (2021) Assessing the effects of delivery attributes on e-shopping consumer behaviour. *Sustainability* 14(1):13.

Duque D, Morton DP, Singh B, Du Z, Pasco R, Meyers LA (2020) Timing social distancing to avert unmanageable COVID-19 hospital surges. *Proc. Natl. Acad. Sci. USA* 117(33):19873–19878.

Etminani-Ghasrodashti R, Hamidi S (2020) Online shopping as a substitute or complement to in-store shopping trips in Iran? *Cities* 103:102768.

Farag S, Krizek KJ, Dijst M (2006) E-shopping and its relationship with in-store shopping: Empirical evidence from the Netherlands and the USA. *Transportation Rev.* 26(1):43–61.

Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, Prill M, et al. (2020) Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019-covid-net, 14 states, March 1–30, 2020. *Morbidity Mortality Weekly Rep.* 69(15):458.

Goel R, Sharma R (2020) Mobility based sir model for pandemics-with case study of COVID-19. *2020 IEEE/ACM Internat. Conf. Adv. Social Networks Analysis Mining* (IEEE, Piscataway, NJ), 110–117.

Grefenstette JJ, Brown ST, Rosenfeld R, DePasse J, Stone NT, Cooley PC, Wheaton WD, et al. (2013) FRED (a framework for reconstructing epidemic dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health* 13(1):1–14.

Hale T, Angrist N, Kira B, Petherick A, Phillips T, Webster S (2020) Variation in government responses to COVID-19, Version 6.0. Working paper, Blavatnik School of Government, Oxford, UK.

Hardy BL, Logan TD (2020) *Racial Economic Inequality Amid the COVID-19 Crisis* (The Hamilton Project, Washington, DC).

Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, Druckenmiller H, et al. (2020) The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* 584(7820):262–267.

Hsiao MH (2009) Shopping mode choice: Physical store shopping vs. e-shopping. *Transportation Res. Part E Logist. Transportation Rev.* 45(1):86–95.

Huang Y, Oppewal H (2006) Why consumers hesitate to shop online: An experimental choice analysis of grocery shopping and the role of delivery fees. *Internat. J. Retail Distribution Management* 34(4/5):334–353.

International Monetary Fund (2020) World economic outlook: A long and difficult ascent. Accessed April 28, 2024, https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020.

International Monetary Fund (2021) World economic outlook: Managing divergent recoveries. Accessed April 28, 2024, https://www.imf.org/en/Publications/WEO/Issues/2021/03/23/world-economic-outlook-april-2021.

Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA (2020) Population flow drives spatio-temporal distribution of Covid-19 in China. *Nature* 582(7812):389–394.

Kim MJ, Lee S (2021) Can stimulus checks boost an economy under COVID-19? Evidence from South Korea. *Internat. Econom. J.* 35(1):1–12.

Koch J, Frommeyer B, Schewe G (2020) Online shopping motives during the COVID-19 pandemic lessons from the crisis. *Sustainability* 12(24):10247.

Kraemer MU, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, du Plessis L, et al. (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368(6490):493–497.

Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, Wesolowski A, et al. (2020) Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 585(7825):410–413.

Le HTK, Carrel AL, Shah H (2022) Impacts of online shopping on travel demand: A systematic review. *Transportation Rev.* 42(3):273–295.

Mantin B, Koo B (2010) Weekend effect in airfare pricing. *J. Air Transportation Management* 16(1):48–50.

Mosteller J, Donthu N, Eroglu S (2014) The fluent online shopping experience. *J. Bus. Res.* 67(11):2486–2493.

Nguyen HV, Tran HX, Van Huy L, Nguyen XN, Do MT, Nguyen N (2020) Online book shopping in Vietnam: The impact of the COVID-19 pandemic situation. *Public Res. Quart.* 36(3):437–445.

Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, Agha M, et al. (2020) The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Internat. J. Surgery* 78:185–193.

Nouvellet P, Bhatia S, Cori A, Ainslie KE, Baguelin M, Bhatt S, Boonyasiri A, et al. (2021) Reduction in mobility and COVID-19 transmission. *Nature Comm.* 12(1):1–9.

Paoluzzi M, Gnan N, Grassi F, Salvetti M, Vanacore N, Crisanti A (2021) A single-agent extension of the sir model describes the impact of mobility restrictions on the COVID-19 epidemic. *Sci. Rep.* 11(1):1–10.

Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS, Hanif W, et al. (2020) Ethnicity and COVID-19: An urgent public health research priority. *Lancet* 395(10234):1421–1422.

Poletto C, Tizzoni M, Colizza V (2013) Human mobility and time spent at destination: Impact on spatial epidemic spreading. *J. Theoretical Biology* 338:41–58.

Putra S, Mutamar ZK (2019) Estimation of parameters in the sir epidemic model using particle swarm optimization. *Amer. J. Math. Comput. Model* 4:83–93.

Roggeveen AL, Sethuraman R (2020) How the COVID-19 pandemic may change the world of retailing. *J. Retailing* 96(2):169.

Salgotra R, Gandomi M, Gandomi AH (2020) Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos Solitons Fractals* 138:109945.

Salomon JA, Reinhart A, Bilinski A, Chua EJ, La Motte-Kerr W, Rönn MM, Reitsma MB, et al. (2021) The US COVID-19 trends and impact survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proc. Natl. Acad. Sci. USA* 118(51):e2111454118.

Sayyida S, Hartini S, Gunawan S, Husin SN (2021) The impact of the COVID-19 pandemic on retail consumer behavior. *Aptisi Trans. Management* 5(1):79–88.

Schmid B, Schmutz S, Axhausen KW (2016) Exploring the choice between in-store and online shopping. *Proc. 23rd Internat. Conf. Recent Adv. Retailing Service Strategies* (IVT, ETH Zurich, Zürich).

Sharma A, Jhamb D (2020) Changing consumer behaviours toward online shopping-an impact of COVID-19. *Acad. Marketing Stud. J.* 24(3):1–10.

Shi K, De Vos J, Yang Y, Li E, Witlox F (2020) Does e-shopping for intangible services attenuate the effect of spatial attributes on travel distance and duration? *Transportation Res. Part A Policy Practice* 141:86–97.

Sood L, Sood V (2021) Being African American and rural: A double jeopardy from COVID-19. *J. Rural Health* 37(1):217.

Texas Medical Association (2020) Risk assessment chart. Accessed January 7, 2022, https://www.texmed.org/uploadedFiles/Current/2016_Public_Health/Infectious_Diseases/309193%20Risk%20Assessment%20Chart%20V2_FINAL.pdf.

Van Dorn A, Cooney RE, Sabin ML (2020) COVID-19 exacerbating inequalities in the US. *Lancet* 395(10232):1243.

Warren MS, Skillman SW (2020) Mobility changes in response to COVID-19. Preprint, submitted March 31, https://arxiv.org/abs/2003.14228.

Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, Chen-Li D, et al. (2020) Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *J. Affective Disorders* 277:55–64.

Xu Z, Glass K, Lau CL, Geard N, Graves P, Clements A (2017) A synthetic population for modelling the dynamics of infectious disease transmission in American Samoa. *Sci. Rep.* 7(1):1–9.

Yancy CW (2020) COVID-19 and African Americans. *JAMA* 323(19):1891–1892.