Data on the Health and Economic Impacts of COVID-19

Meta Brown, Hanbat Jeong, Kurt Lavetti, and Bruce Weinberg[1] (Lead), Ohio State University

Effective policy responses to a public health crisis require quantifying and convincingly communicating the health and socioeconomic costs and benefits of these policies. Throughout the pandemic, a point of contention has been the perception among large segments of the public that the socioeconomic costs of policy responses, including job loss and isolation, were "worse than the problem itself" (Trump 2020). While we now have at least 26models of COVID-19 infections (https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/forecasts-cases.html), to the best of our knowledge our own work is the first rigorous model that also accounts for the effects on the economy and the economy's effect on health. Yet understanding how social distancing policies affect the economy and in turn feedback to health is essential for understanding the full costs and benefits of the pandemic. Unfortunately, much of the data to rigorously assess and quantify these effects simply do not exist. For instance, data are only available on confirmed cases, while actual cases, which are essential for understanding the full impact and risk of infection, are latent. Second, understanding disparities requires rich data by age, gender, race, and ethnicity, but such breakdowns are not available. This project will use a frontier-expanding model to impute rich data on actual infections, hospitalization, mortality, and vaccination by gender, race, ethnicity, and age for U.S. counties on a weekly basis. These imputed series will first inform the discourse and second allow our team and other researchers to quantify the health and socioeconomic costs and benefits of policy responses to COVID-19 rigorously and neutrally, simulate counterfactuals, and estimate the implied tradeoffs to date and going forward.

Under current, short-run supplements from NIA through NBER and from NCATS, we have rapidly built, estimated, and simulated a frontier-expanding model that includes many key features of the COVID-19 pandemic. Specifically, we have estimated a model, sketched in Figure 1, that allows for the geospatial spread of COVID-19 infections that embeds policy responses and bi-directional interactions between health and economic outcomes. Two additional key features of the COVID-19 pandemic and other public health crises are that the incidence (here the number of COVID-19 infections) is latent or, more colloquially, "misunderestimated" and that there are greatly disparate impacts across groups. Here we propose to extend our existing model to allow for latent infection rates and disparities across groups. To do so, we both require and will generate high-frequency, county-level imputations of infections for detailed specific groups. Such estimates will allow us to estimate how, for instance, actual infection levels in one group (e.g. a particular gender, race / ethnicity, age group) spillover to others in the same and other groups in the same and other locations and, for instance, group-specific COVID-19 mortality rates. These imputations will be validated against published statistics (e.g. more aggregated and lower frequency data) and distributed to the broader research community.

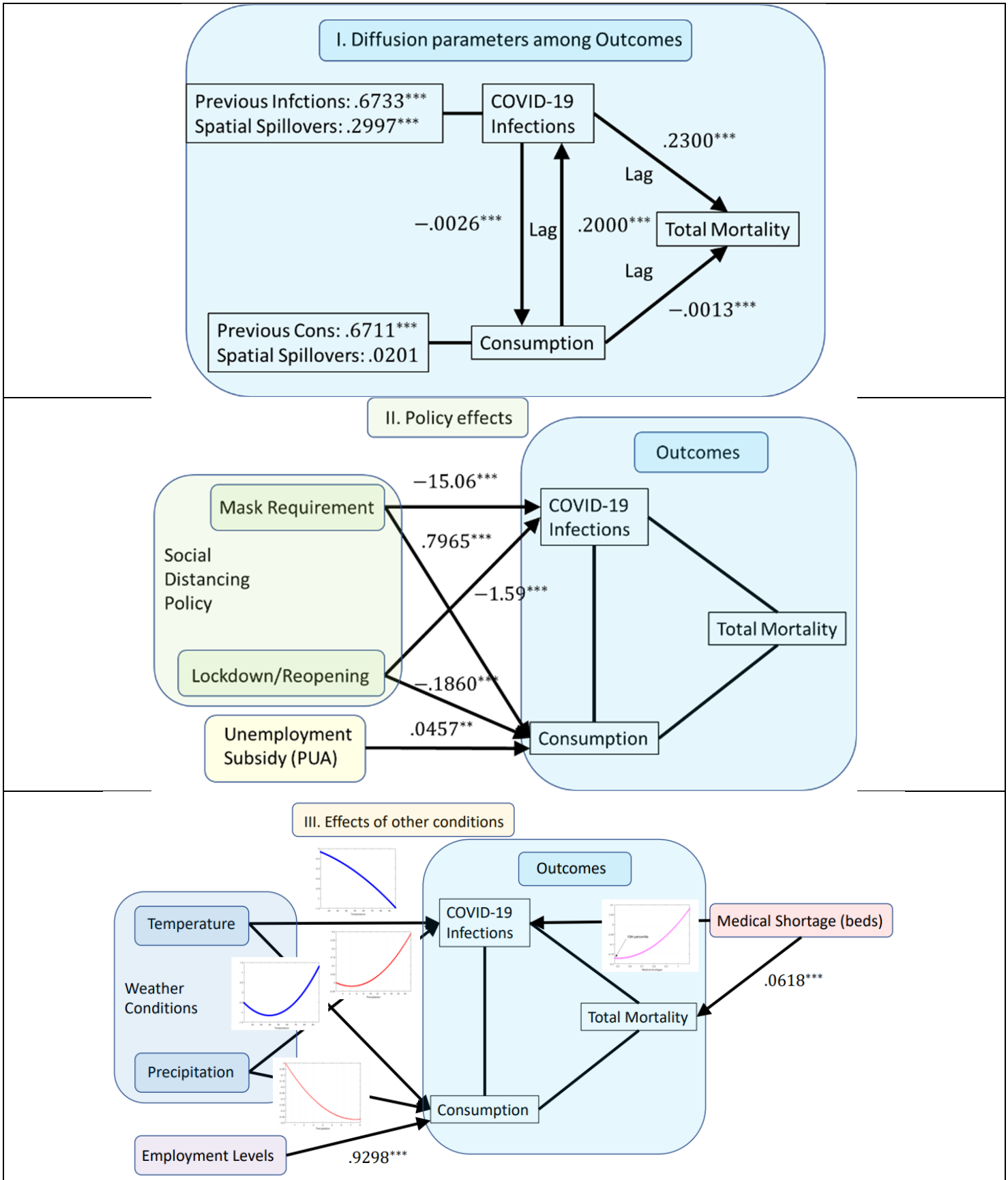
Formally, our model consists of one equation for each of our outcomes: latent COVID-19 infections, economic activity, and (all-cause) mortality. We will also include "observation equations" that give group-specific confirmed cases and vaccinations as a function of latent infections and vaccinations and measurement errors. The observation equation is modeled using a Kalman (1960) filter and estimated iteratively by solving Ricatti (1724) equations. Currently, we are using weekly, count-level data through November of 2020 but have data in hand for most of 2021. Infections and mortality in each county-week-demographic group are modeled using a spatial dynamic panel data (SDPD) model (Shi and Lee, 2017; and Jeong and Lee, 2020), where latent infections among a group in a county in one week depend on infections among all groups in that and nearby counties in the previous week as well as previous economic activity. Economic activity in a county in one week depends on economic activity in that and nearby counties in previous weeks and latent infections in that county in previous weeks. We use a control function approach to estimate causal effects (Heckman and Robb, 1985). Intuitively, the control function approach applies omitted variable logic to model the endogenous portion of the errors in the main equations in auxiliary equations to control for the portion of the policies that are correlated with omitted factors. We currently have compelling estimates in hand for the model that does not include latent infections, rich demographic groups, or vaccination.

[1] Address correspondence to: weinberg.27@osu.edu.

We will augment existing data using two sources. First, the U.S. Centers for Disease Control and Prevention (CDC) provides record-level information on COVID-19 status (e.g., confirmed or probable cases) and outcomes (e.g. hospitalization and mortality) with data on residence, date, age, gender, race / ethnicity. We will aggregate these data to provide weekly, county-level estimates that will enter our model. As indicated, we will use a Kalman filter approach to generate estimates of actual, latent infections by county, week, and demographic group that will be redistributed. Second, the Kaiser Family Foundation (KFF) provides weekly state-level vaccinated populations by race and ethnicity, but not age or gender. Rich data on vaccinations is a key explanatory variables for COVID-19 infections. We will use a multi-step procedure to generate rich, time-varying, county-level, group-specific data. The first step is to interpolate (e.g., using a moving average) to eliminate missing values for specific periods. The second step is to impute county-level vaccinated populations across groups from the two data sources given that regular, county-level data on vaccinated populations by age and gender are not available. By assuming that a vaccinated person is not affected by COVID-19 (or some specified immunity rate from the medical literature), we suppose the existence of a functional mapping between regional COVID-19 infections for a specific group and its vaccination rate. The third step is to recover this mapping by matching the state-level race / ethnicity data from KFF to the corresponding COVID-19 infections across race / ethnicity groups from CDC. By employing this functional mapping, we impute various groups' vaccinated populations for each county and time period. These estimates can be generated from our model or separately and used within our model.

Once estimated, our model estimates will be of direct interest. They will show how infections propagate across geographies and demographic groups. Our estimates will also generate rich, imputed series on actual infections, hospitalization, mortality, and vaccination rates overall and by demographic group for each county and week that will be validated and made available to other researchers. Our model can use be used to simulate policy counterfactuals. Specifically, we have designed it to simulate an infections-economic activity possibilities frontier for each county that traces out the lowest feasible infection levels for each level of economic activity from a given set of initial conditions. Our current estimates, shown in Figure 1, imply that masks are a win-win, reducing infections while actually increasing economic activity, but that lockdowns are a win-lose, reducing infections at the expense of reducing economic activity. Thus, it implies that the optimal policy is to have universal mask requirements, but a moderate level of lockdowns. The estimates also indicate very high persistence of infections and high persistence of economic activity, implying strong path dependence. Under the proposed work, we can simulate not only aggregate effects, but the effects on specific demographic groups and the effects of vaccination. For instance, we can simulate how increased vaccination in a county for one group will affect infections and mortality among that and other groups and economic activity in that and other counties. Our intuition is that vaccination will act as a win-win, much like masks, reducing infections and increasing economic activity, but rigorously quantifying the magnitudes and the disparate impacts across groups is essential.

Figure 1. Model Structure and Estimates from our Model without Measurement Error or Vaccines.



Note: Numbers shows the percentage change when each variable increases by 1%. * indicates a 10% level of significance; ** 5% level of significance; and *** 1% level of significance.

REFERENCES

Heckman, J. J., and Robb, R. (1985). “Alternative methods for evaluating the impact of interventions” *Journal of Econometrics* 30 (1–2): 239–267.

Jeong, Hanbat and Lung-Fei Lee (2020). “Spatial dynamic model with intertemporal optimization: specification and estimation” *Journal of Econometrics* 218(1), 82-104.

Kalman, Rudolf E. (1960). “A new approach to linear filtering and prediction problems” *Journal of Basic Engineering* 82, 35-45.

Riccati, Jacopo (1724). “Animadversiones in aequationes differentiales secundi gradus (Observations regarding differential equations of the second order)” *Supplementa* 8, 66-73.

Shi, W., and L.F. Lee, (2017). “Spatial dynamic panel data models with interactive fixed effects” *Journal of Econometrics* 197, 323-347.

Trump, Donald J. (2020). @realDonaldTrump “WE CANNOT LET THE CURE BE WORSE THAN THE PROBLEM ITSELF.” *Twitter,* 22 Mar. 2020, 11:50 p.m., https://twitter.com/realDonaldTrump/status/1241935285916782593.