Note: Comments and discussion cover two papers presented at the Session 4 of [Summer 2020 BPEA](#) conference on labor markets and the economics of non-pharmaceutical interventions.

Comments and Discussion

SUMMARY OF COMMENT

VICTOR CHERNOZHUKOV provided oral comments. He congratulated the authors on providing such rapid and innovative data on economic activity early in the pandemic.

His comments focused on some of the challenges of estimating the effect of non-pharmaceutical interventions (NPIs) on economic activity. He presented weekly correlations between seven distinct NPIs (state-level data, March through May 2020). Nearly all the correlations exceeded 0.8, and several exceeded 0.9, indicating scope for omitted variable bias in regressions by Gupta, Simon, and Wing and by Bartik and colleagues, which considered only a subset of NPIs. Another econometric challenge is that the policies considered in these data were "hard" policies that took effect at a specific known date, while policies that changed behavior more gradually were excluded. Policies that induce gradual behavioral change, if not measured and included, would induce patterns that these regressions could misattribute as endogenous self-protection. As an example, Chernozhukov turned to some of his own research with Hiro Kasahara and Paul Schrimpf on use of masks.[1] They found a large effect of masking orders on cases, deaths, and mobility, both through a direct channel and through a behavioral channel. These and other econometric considerations led him to speculate that both papers—by Gupta, Simon, and Wing and Bartik and colleagues—could underestimate the effect of policies on economic activity.

1. Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf, "Causal Impact of Masks, Policies, Behavior on Early COVID-19 Pandemic in the U.S.," *Journal of Econometrics* 220, no. 1 (2021): 23–62.

COMMENT BY

CAROLINE BUCKEE The deadly COVID-19 pandemic emerged in early 2020 and, in the absence of effective treatments or a vaccine, led to the unprecedented implementation of socially and economically disruptive non-pharmaceutical interventions around the world. In the two papers by Bartik and colleagues and by Gupta, Simon, and Wing the impact of these interventions on employment and human behavior, respectively, are examined, and in both papers, the authors use data streams from mobile phones to measure social and economic activity in relation to the dynamics of the labor force and public health policies around the United States. The comments below reflect my background as an infectious disease epidemiologist and as a researcher who has been using mobile phone data to monitor movement patterns in the context of disease modeling for nearly a decade. I have focused on two aspects that are relevant to both studies: the importance of spatially heterogeneous disease burden and the use of mobile phone data as a proxy for human behavior.

THE IMPORTANCE OF SPATIAL HETEROGENEITIES IN THE BURDEN OF COVID-19 Both studies examine economic and behavioral time series data in relation to policies that were implemented to slow the transmission of SARS-CoV-2. As they find, and as others have observed (Badr and others 2020), people across the country reacted strongly to the declaration of a national emergency on March 13 regardless of local policies. Almost any measure of mobility or other behavior is likely to show this rapid countrywide decline in activity in response to the threat of the pandemic. Most analyses, including these two, have concluded that the synchronization of behavior may have resulted from individuals acting based on national and global information about the pandemic rather than local policies. Indeed, Bartik and colleagues note that their results with respect to labor markets and economic activity "have more to do with broader health and economic concerns affecting product demand and labor supply" than with the timing of specific policies.

However, the trajectory of the epidemic in the United States has been characterized by distinct geographic heterogeneities within and between individual states, among different demographics, and even within cities (Kissler and others 2020). These heterogeneities reflect the spatial progression of the epidemic across the country, starting in Seattle and New York before moving into the south and middle of the country over the summer, as well as remarkable local heterogeneities resulting from income and racial inequalities. Both of these types of heterogeneity have implications for the interpretation of economic and mobility data because decision making by individuals generating the data reflect very different experiences of the disease itself.

Although people's behaviors in response to the national lockdowns were relatively synchronized across the country, their perceptions of the risks posed from COVID-19 are likely to have been strongly dependent on their personal, local experiences. People in New York may have experienced illness or death among friends and loved ones or witnessed the fatigue and desperation of health workers in their communities. In contrast, recent seroprevalence estimates suggest that even by June, much of the Midwest had not yet experienced any significant SARS-CoV-2 transmission (Anand and others 2020). Not only would this have an impact on individuals' real and perceived risks from COVID-19 but also on their sense that the economic and social hardships experienced as a result of interventions were justified. To the extent that compliance and reaction to non-pharmaceutical interventions will depend on perceived risks, as we have seen in the context of Ebola in West Africa (Peak and others 2018), many of the nationwide metrics analyzed in these studies may mask significant regional heterogeneity. In particular, the speed and behavioral response to reopening, including consumer behavior, people leaving home and mixing socially, and the likelihood that individuals look for work and re-open their businesses, may have shown significant regional variation.

The second important spatial heterogeneity in disease incidence and burden is highly local and reflects structural disparities between neighborhoods that fall along socioeconomic and racial lines. Indeed, Bartik and colleagues find significant differences in employment and rehiring between different racial groups and income levels. Just as regional differences in disease burden may have had an impact on state-level economic and behavioral metrics, local differences in the experience of disease and death from COVID-19 are likely to have been pronounced among these economic categories. Consistent with nationwide racial disparities in mortality due to COVID-19 (Bassett, Chen, and Krieger 2020), analyses of COVID-19 deaths in Cook County, Illinois, found startling mortality rate differences due to COVID-19 between neighborhoods depending on poverty and race, varying from 14.1 per 100,000 in wealthy neighborhoods among white people, to 135.1 per 100,000 in poor neighborhoods among Hispanic and Latinx people (Feldman and Bassett 2020; Acosta and Irizarry 2020). A seroprevalence study among pregnant women in New York City in April showed a cumulative incidence of 11 percent in Manhattan versus 26 percent in South Queens, for example (Kissler and others 2020). In that study we showed that local differences in commuting behavior, measured using mobility data from Facebook users, was strongly correlated with seroprevalence. Thus, both mobility behavior related to employment and COVID-19-related illness and death have had an impact even on people living in the same city differently.

Studies aiming to understand social and economic decisions made by individuals in relation to public health and other policies—as both studies presented here seek to do in different ways—may therefore gain important insights if they account for the dramatic differences between individuals in their local experience of the epidemic when interventions were imposed or lifted.

THE USE OF MOBILITY DATA FROM PRIVATE COMPANIES AS A PROXY FOR HUMAN ACTIVITY Both Bartik and colleagues and Gupta, Simon, and Wing derive quantitative behavioral estimates from SafeGraph data, and Gupta, Simon, and Wing go further and use multiple different sources of activity data (for example, from Google and Apple) from mobile phones. Gupta, Simon, and Wing note that while mobility data from mobile phones have become relatively routine among infectious disease epidemiologists, they are still quite rare in other fields. While mobile phone data are a useful nearly real-time proxy for human behaviors, including for monitoring human behavior during this pandemic, there are a number of important issues that—in my opinion—make it challenging to directly use derived metrics in a quantitative, statistical analysis.

Gupta, Simon, and Wing discuss some of these caveats, including the representativeness of the data with respect to demographic structure, but it is important to outline some of the other systematic biases that may have an impact on analyses. These have been reviewed in the context of COVID-19 in Grantz and others (2020) and Oliver and others (2020), and a standardization of mobility metrics of this kind has been called for (Kishore and others 2020).

So-called ad tech data, such as the data from SafeGraph, can be distinguished from other data sources, including Google, Apple, Facebook, or data from mobile operators. Ad tech data derive from advertisements associated with the use of particular apps on smart phones, and the data from individuals are processed and packaged by multiple companies before they are analyzed. This creates opacity around the biases and details of individual data sets, including missingness, and data imputation or inference is often performed prior to release of the data. Therefore, even an investigation of the biases in the data becomes impossible for research groups using the data. Indeed, unlike data from Facebook, for example, where data quality or missingness is sometimes reported, this imputation step means that uncertainty in the SafeGraph estimates is impossible to ascertain.

Demographic biases are clearly an issue, because most mobility data from mobile phones reflect smart phone users only, who skew young and wealthy (mobile operator data are an exception because they include "dumb phone" subscribers, which is why operator data are often more appropriate in low-income settings). With respect to representativeness, unlike Google or Facebook, ad tech data providers often report their "monthly active users" (MAU), but this can be misleading. For example, 1 million monthly active users is not the same as a longitudinal sample of 1 million individuals because a user may appear infrequently or only once in the data set, and the number of users can vary dramatically from day to day. This high turnover is rarely reported, making it difficult to quantify uncertainty associated with any particular day and location. There are, in addition, geographic variations in representativeness that cannot be accounted for. For example, by comparing Facebook data to SafeGraph data across the United States, we find that while Facebook reports missingness in rural counties, SafeGraph imputes data and reports no missingness (personal communication).

Demographic and geographic representativeness aside, mobility metrics derived from these data sets—such as the mixing index used by Gupta, Simon, and Wing—are difficult to interpret. Standardized analytical frameworks, particularly validated ones, are still absent for this kind of data (Kishore and others 2020). Interpreting mixing indexes and other metrics of mobility is also complicated by the fact that in a large, geographically diverse country, the same movement patterns may represent very different behaviors in urban versus rural locations. Out-of-county travel, for example, is hard to interpret in the absence of spatial context, even when compared to a baseline, because it may depend on the spatial layout of grocery stores and so on. Gupta, Simon, and Wing include multiple metrics and data sources as a way to confirm their findings, which makes sense, but since all the metrics are likely to be biased in the same ways (reflecting smart phone users) there may still be bias unaccounted for. Taken together, although the qualitative findings are important and useful, these issues with uncertainty about data quality and representativeness and the rigor of particular derived metrics mean that making sense of effect sizes from time series and statistical analyses is challenging.

CONCLUSIONS Both studies track the behavioral and economic impacts of the unprecedented public health interventions that were put in place due to COVID-19 earlier this year. As we move into autumn and face a long winter with possible renewal of various behavioral interventions, understanding how people and the economy will respond is critical. Mobile phone data are a valuable source of information about human activity, although they are a loose proxy for the contacts that spread the virus and likely to be increasingly difficult to interpret epidemiologically against the backdrop of layered interventions such as masking. I don't necessarily expect the reaction to future lockdowns to recapitulate the behavioral dynamics we saw in the spring, not only because the economic and political situation is different now, but also, crucially, because now there are hardly any US communities that have not suffered significant illness and death due to COVID-19, and this will change the social and political acceptability of interventions.

REFERENCES FOR THE BUCKEE COMMENT

Acosta, Rolando J., and Rafael A. Irizarry. 2020. "Monitoring Health Systems by Estimating Excess Mortality." MedRxiv. https://www.medrxiv.org/content/10.1101/2020.06.06.20120857v2.

Anand Shuchi, Maria Montez-Rath, Jialin Han, Julie Bozeman, Russell Kerschmann, Paul Beyer, Julie Parsonnet, and Glenn M. Chertow. 2020. "Prevalence of SARS-CoV-2 Antibodies in a Large Nationwide Sample of Patients on Dialysis in the USA: A Cross-Sectional Study." *The Lancet* 396, no. 10259: 1335–44. https://doi.org/10.1016/S0140-6736(20)32009-2.

Badr, Hamada S., Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M. Squire, and Lauren M. Gardner. 2020. "Association between Mobility Patterns and COVID-19 Transmission in the USA: A Mathematical Modelling Study." *The Lancet Infectious Diseases* 20, no. 11: 1247–54. https://doi.org/10.1016/S1473-3099(20)30553-3.

Bassett, Mary T., Jarvis T. Chen, and Nancy Krieger. 2020. "Variation in Racial/Ethnic Disparities in COVID-19 Mortality by Age in the United States: A Cross-Sectional Study." *PLoS Medicine* 17, no. 10: e1003402. https://doi.org/10.1371/journal.pmed.1003402.

Feldman, Justin M., and Mary T. Bassett. 2020. "The Relationship between Neighborhood Poverty and COVID-19 Mortality within Racial/Ethnic Groups (Cook County, Illinois)." MedRxiv. https://www.medrxiv.org/content/10.1101/2020.10.04.20206318v1.

Grantz, Kyra H., Hannah R. Meredith, Derek A. Cummings, C. Jessica E. Metcalf, Bryan T. Grenfell, John R. Giles, and others. 2020. "The Use of Mobile Phone Data to Inform Analysis of COVID-19 Pandemic Epidemiology." *Nature Communications* 11, article no. 4961. https://doi.org/10.1038/s41467-020-18190-5.

Kishore, Nishant, Mathew V. Kiang, Kenth Engø-Monsen, Navin Vembar, Andrew Schroeder, Satchit Balsari, and Caroline Buckee. 2020. "Measuring Mobility to Monitor Travel and Physical Distancing Interventions: A Common Framework for Mobile Phone Data Analysis." *The Lancet Digital Health* 2, no. 11: e622-e628. https://doi.org/10.1016/S2589-7500(20)30193-X.

Kissler, Stephen M., Nishant Kishore, Malavika Prabhu, Dena Goffman, Yaakov Beilin, Ruth Landau, and others. 2020. "Reductions in Commuting Mobility Correlate with Geographic Differences in SARS-CoV-2 Prevalence in New York City." *Nature Communications* 11, article no. 4674. https://doi.org/10.1038/s41467-020-18271-5.

Oliver, Nuria, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Deletaille, Marco De Nadai, and others. 2020. "Mobile Phone Data for Informing Public Health Actions across the COVID-19 Pandemic Life Cycle." *Science Advances* 6, no. 23: eabc0764. https://doi.org/10.1126/sciadv.abc0764.

Peak, Corey M., Amy Wesolowski, Elisabeth zu Erbach-Schoenberg, Andrew J. Tatem, Erik Wetter, Xin Lu, and others. 2018. "Population Mobility Reductions Associated with Travel Restrictions during the Ebola Epidemic in Sierra Leone: Use of Mobile Phone Data." *International Journal of Epidemiology* 47, no. 5: 1562–70. https://doi.org/10.1093/ije/dyy095.

GENERAL DISCUSSION Jason Furman inquired about the nature of job loss over time. Furman remarked that it is possible that if weekly unemployment insurance (UI) claims remain high throughout the summer, then those unemployment spells may be different in nature. For example, he posited that some initial job losses could be primary, direct effects of the COVID-19 pandemic but that it is possible subsequent job losses could be the result of more traditional recession forces. Furman speculated that by determining this distinction between types of job loss, policymakers may be able to gain insights into how and when those jobs might be recovered.

Hilary Hoynes speculated whether it would be possible to link the private sector Homebase data used in the paper with recently published data from the Treasury Department on the Paycheck Protection Program (PPP). Hoynes suggested it would be interesting to see if there could be a way to see to what extent the PPP affected labor market outcomes for workers in the Homebase data. More specifically, she wondered whether such a linking could shed light on whether PPP loans accomplished certain goals policymakers had for it (e.g., keeping workers connected to their employers).

Adding to this conversation, Marianne Bertrand pointed out that the Treasury Department plans to release detailed data on the name of firms, location, firm size, and so on, for the larger loans (above $150,000). She pointed out that when these data become available, it could be possible to link the firms in the Treasury Department's PPP loan data with the firms in the Homebase data set.

Simon Mongey shared a resource from the Philadelphia Federal Reserve Bank on the PPP loans.[1]

Stephen Goss asked the discussant Caroline Buckee about the effects of seasonality and weather on the spread of the coronavirus. He mentioned that some observers have pointed to Brazil, which, being in the Southern Hemisphere and currently in the midst of winter, has still seen a surge in cases. Goss inquired whether Brazil's experience might provide insights into what sort of experience the United States and the European Union (EU) may have with the virus as our seasons begin to change. He speculated whether the EU's current relative success in controlling the virus may be short-lived as the weather begins to change.

Henry Aaron asked whether improved treatment methods are being incorporated into models. He remarked that it seems much of the conversation has surrounded spread and deaths but not much on changes in treatment.

In response to Goss's comments, Buckee says that because other coronaviruses do exhibit seasonal effects, it is likely that this strand may be affected by seasonality, but to a limited degree. The much more relevant way that seasonality will play a role is in the gathering of people indoors as a result of the colder weather in the fall and winter months. Buckee worried about the potential surge in cases that may result if many of the social interactions that have occurred outdoors during the summer continue indoors in the fall. In particular, she was concerned about schools reopening in the fall without the proper precautions being taken. As for the comparison between the United States and the EU, Buckee argued that the difference in success with dealing with the virus has largely been an effect of policy choices: lack of increased testing capacity, issues surrounding social response and messaging, and so on.

Addressing Aaron's question, Buckee replied that changes in treatment methods have not shown through in the data, largely because there have not been many significant breakthroughs in treatments. In addition to the many ongoing trials, Buckee referred specifically to a recent trial of dexamethasone that showed a 30 percent reduction in deaths among people on ventilators. However, she pointed out that many of those trial results haven't been

1. "SBA_PPP," public data tables on the Payroll Protection Program by the US Small Business Administration, GitHub, https://github.com/RocArm/SBA_PPP?fbclid=IwAR0hHw_lJObIzRoroYWLhWU7RcpiXDszdIkdsMCRz3VLKNMQ4GSsngUwBw.

rolled out widely yet, which is why she didn't think that these trials were having a major impact on treatment and the death rate. A related point that Buckee made in this conversation was that a large share of deaths early on in the pandemic occurred in nursing homes and assisted care facilities. More recently, as states have begun reopening, the largest surge in cases has been among young people, who have a lower mortality rate anyway.[2] In light of these two trends, Buckee commented that it's difficult to disentangle whether that change reflects a demographic shift, differences in social distancing behavior, or household structure differences in different geographic areas as the epidemic spreads across the country. Buckee concluded that while it can be hard to discern exactly what's happening, these trends will be important moving forward.

Austan Goolsbee highlighted a recent paper that he and Chad Syverson have put out that uses county-level lockdown policies (rather than state-level policies).[3] Goolsbee claimed that their paper finds that looking at county-level policies as opposed to state-level ones seems to matter a fair amount: many of the hardest-hit counties implemented policies well before their states did. Goolsbee mentioned that by doing a horse race on the two levels of policy, they find that the local level appears to be far more influential. He concluded by saying that he and his coauthor have posted the data publicly for anyone to use.

Alessandro Rebucci pointed out he has a paper where he and his coauthors analyze the relationship between partisanship and state-level heterogeneity in compliance with non-pharmaceutical interventions (NPIs).[4] Their empirical evidence shows that preferences and attitudes toward "free" interactions are an additional factor in the decision problem. Responding to this point, Bertrand commented that it would be interesting to think about heterogeneity of order effects between Democratic versus Republican states. She speculated that one can imagine that a truly enforced order in a Republican state may matter more than in a Democratic state if people in

2. Julie Bosman and Sarah Mervosh, "As Virus Surges, Younger People Account for 'Disturbing' Number of Cases," *New York Times*, June 25, 2020, https://www.nytimes.com/2020/06/25/us/coronavirus-cases-young-people.html.

3. Austan Goolsbee and Chad Syverson, "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020," working paper 27432 (Cambridge, Mass.: National Bureau of Economic Research, 2020).

4. Alexander Chudik, M. Hashem Pesaran, and Alessandro Rebucci, "Voluntary and Mandatory Social Distancing: Evidence on COVID-19 Exposure Rates from Chinese Provinces and Selected Countries," research paper 20-03 (Baltimore: Johns Hopkins Carey Business School, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3576703.

Democratic states take the disease more seriously and are adjusting their behavior even absent an order to do so.

Jesse Rothstein thanked both of the discussants and the participants for their helpful comments. Responding to Furman's comment, Rothstein mentioned that the data used in the paper did not allow for that distinction to be drawn, but he pointed to Till von Watcher's recent paper analyzing California UI claims data.[5] Rothstein mentioned that Hedin, Schnorr, and von Wachter find that the first waves of UI claims were concentrated among workers with less educational attainment and workers in specific industries and that subsequent waves of UI claims tended to be more representative of the broader labor force, potentially supporting Furman's hypothesis.

Alexander Bartik echoed Rothstein's thanks and responded to a few participants' points in particular. Building on Rothstein's response to Furman, Bartik highlighted the figure in their paper that shows payroll employment by month and industry. He emphasized that this figure showed that in the early weeks of the pandemic, the leisure and hospitality industry in particular was hard hit; the data at that time had not shown a spread to certain industries (e.g., durable goods, manufacturing, construction, etc.). However, Bartik acknowledged that the data may change in the coming months.

Bartik responded to Hoynes by stating that currently that sort of linking is not yet possible, but that he and his coauthors are working with scholars at Harvard to conduct a survey of the firms in the Homebase data to see if they can use quasi-experimental methods to accomplish a similar goal with regards the PPP loan data. He also pointed to work being done by Granja and others, who have looked into PPP disbursement and employment effects.[6]

Bartik acknowledged that several participants raised the issue of the paper's focus being only on the shutdown orders. He said that they did this for a variety of reasons but that they plan to incorporate the fuller set of policies into future analysis. Given the nature of Homebase data, he pointed out that they should be able to analyze relatively fine measures of timing. He wanted to clarify that they are not taking a strong stance on

5. Thomas J. Hedin, Geoffrey Schnorr, and Till von Wachter, "California Unemployment Insurance Claims during the COVID-19 Pandemic," policy brief (Los Angeles: California Policy Lab, 2020).

6. Joao Granja, Christos Makridis, Constantine Yannelis, and Eric Zwick, "Did the Paycheck Protection Program Hit the Target?," working paper, April 26, 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3585258.

the time effects of information per se, but that their interpretation of those effects was that they reflected reduced consumer demands for in-person services. Bartik pointed out that this reduced consumer demand could, in part, be a function of schools being closed, since school closings change how parents consume in-person services.

Lastly, Bartik commented that although they had not done it yet, it is possible for them to look at their Homebase sample for 2018 and 2019, which could bolster their analysis.

Sumedha Gupta also thanked all of the participants and said that she greatly appreciated their feedback. Responding to Buckee's comments, Gupta acknowledged that she agreed with many of her points, especially regarding the heterogeneity of the data sources. She pointed out that their paper addresses many of these differences in the data sets, which is why they chose to look at all of them in an effort to capture the whole story and to see if that story is consistent. Since, thus far, much of the data they have looked at have been consistent, Gupta felt confident in claiming the direction (even if not the magnitude) of the effect. Gupta also pointed out that some of their analysis did look at some local (rural versus urban) differences. She also highlighted that their analysis found interesting differences when looking at indoor versus outdoor activity.

Responding to Victor Chernozhukov's comments, Gupta expressed interest in learning more about his bias correction approach and stated that she intended to look into some of the papers he recommended to see if they can implement it.

Gupta also acknowledged that there is difficulty in parsing out the timing differences between the state of emergency declarations versus stay-at-home orders, especially since it all happened in about a three-week period. Furthermore, Gupta posited that although it can be quite difficult to disentangle the effects of each of the public policies, she and her coauthors think of the emergency declaration as a sort of "reduced form" effect for several of the other policies; in other words, it is almost as if the emergency declarations triggered the start of many of the other policies. However, Gupta still recognized the importance of doing estimations by including controls for the different policies as well as the need to have linearized, event time studies to see the effects for all of the policies simultaneously.

Gupta concluded by stating that their main takeaway is that while there has clearly been a policy response (regardless of how wide-ranging the policies one chooses to include), their data seem to suggest the larger effect has been a private response to this pandemic.