View: Indian policymakers need to perfect timing to maximize impact

Synopsis

Taking such decisions requires information - information that is accurate, frequent and actionable. The second wave of Covid in India earlier this year stretched resources to the brink. In such a situation, knowing beforehand where to target these resources can lead to saving of many lives. Waiting for 'where' and 'when' based on delayed information can fuel the spread, rather than stem it.

The last 20 months may have been some of the most challenging for policymakers across the world. Policymaking in, and for, a pandemic is challenging. Time of response and timing of response are both of essence. As highlighted by Narendra Modi's chairing a high-level meeting on Saturday on the new Omicron variant of coronavirus, and India's current vaccination situation, the situation on the ground changes almost daily. Thus policy responses have to be in that timeframe too.

Similarly, the timing of the response is critical too. Take the case of mobility restrictions. Impose them too soon and you adversely affect the economy. Impose them too late and risk losing more lives and people affected by severe infections, in turn adversely affecting the economy. Timing, in fact, is key.

Taking such decisions requires information - information that is accurate,

frequent and actionable. The second wave of **Covid** in India earlier this year stretched resources to the brink. In such a situation, knowing beforehand where to target these resources can lead to saving of many lives. Waiting for 'where' and 'when' based on delayed information can fuel the spread, rather than stem it. This is where 'real-time data' comes in. De-identified and aggregated datasets based on underlying transactional **data** can provide critical insight into these questions.

India Development Foundation research (bit.ly/3HZ285c) published in October 2021 shows that mobility data from Facebook can be used to provide early warning about Covid spread. In a different use case, these data can be used to identify mass scale distress movements, like the ones that saw with migrants moving back home in the nationwide lockdown last year.

De-Identity Kit

The use of such data elsewhere in the world is considerably more advanced. Opportunity Insight's economic recovery tracker (tracktherecovery.org) uses real-time data from multiple sources - both government and private - to track geographically disaggregated economic recovery in the US. The data used in the tracker includes data from credit and debit card companies to get information on

consumer spending and small business activity, job postings, payroll data, participation in online learning activity, mobility data as well as data from the US department of labour and other government sources. In India, too, the source data to construct many of these series exists.

The brilliant thing about these datasets is that they are the byproducts of already existing business or government activity. Carrying out a survey to get information has both upfront and variable costs. With de-identified and aggregated data there is only the upfront cost - designing the datasets, planning and setting up the algorithms to de-identify and aggregate the data. The variable cost is close to zero. And, thus, the frequency of such data is decided not by the cost but by the frequency of the underlying transactions. This is a boon for policymaking.

A key challenge in creating these datasets is conceptual and technological de-identification of the data. For the digital economy to function, it is important for consumers to trust that their data is safe and secure, and is not being used in ways that may not be desirable to them. These publicly available datasets - such as the Facebook mobility data - typically use conceptual techniques such as masking, and computing techniques such as differential privacy algorithms, to ensure privacy.

These techniques ensure that data pertaining to one individual or small group cannot be reconstructed from the aggregated data. They, thus, allow the de-identification of data in a way that maintains the privacy of individual users while allowing the dataset to be used for an important public purpose.

Dismiss At Our Own Risk

Among researchers there is a tendency to dismiss the use of these novel datasets.

since they are not representative of the population, especially in a context like India where digital penetration is still relatively low. Representativeness here means that the data capture the information from all possible groups in society. This is asking the wrong questions and then dismissing the data, since it can't answer these questions. These data can be extremely helpful in understanding the direction of change.

Are grocery stores seeing the same number of transactions now than they did before the pandemic? Do fewer people move between Mumbai and Pune when Covid increases than otherwise? And then the next step. Why have transactions changed? Is it demand recovery, or more cashless transactions? Do people move less voluntarily, or because of external restrictions? For a host of policy decisions, it is knowledge about the change that really matters.

The source data to construct the various series used in the economic recovery tracker mentioned earlier also exists in India. So why don't the datasets themselves exist? Why are they not being worked on by researchers to provide the all-needed critical insight to policymaking? And this applies both to data from government and private sources.

Lack of access to data is a common theme in policy discourse in India. But is the real constraint actually the availability of data? Or rather the availability of technological skills, insight and an enabling environment that provides the incentives to innovate with data? What really are the critical factors in short supply here?