Interstate migration and spread of Covid-19 in Indian states

Debajit Jha[1] and Anirban Chakraborti[2]

Abstract: There has been a huge impact of the Covid-19 pandemic on the Indian economy. Almost all the sectors of the economy have been severely impacted. The vulnerability of the economy was highest after the imposition of the nationwide lockdown in March 2020, when thousands of migrant workers were forced to return to their homes from the large or metropolitan cities of the country. After the lockdown was lifted and some months passed, people had started migrating again to their work places. However, at the advent of the second wave and state-wise lockdowns imposed, a similar exodus of migrants was seen. While this humanitarian crisis and the associated government policies in India have been heavily discussed in the national and international policy discourses, there has been no systematic study or significant estimates of the mobility of migrants during the pandemic. Although reverse migration has been perceived to be one of the primary reasons for the spread of the virus in the Indian states, the unavailability of a consistent dataset on reverse migration has severely limited academic research on the relationship between reverse migration and the spread of Covid-19 in the country. This has further affected policymaking for tackling the crisis.

In this paper, first, we have tried to construct a *proxy* dataset of "reverse migration" using interstate train running information during the second wave of the Covid-19 pandemic. To understand the complex trend and pattern of interstate human mobility during the second wave of the Covid-19 pandemic, we construct the complex networks of mobility of people using the train running data. Then, we have carried out an econometric exercise to identify the relationship between reverse migration and the spread of Covid-19 in different Indian states, using both the train running information data that we have constructed, as well as the Census-2011 data on migration. The results show that both train migration and census migration have positive and significant impact on the spread of the Covid-19 cases in the Indian states. Besides migration, population density per square kilometre, percentage of the urban population, and state per capita income were also found to have a positive impact on the spread of the virus. Moreover, we found that the availability of hospital beds in different states has helped to reduce the spread of the virus.

Keywords: Covid-19, pandemic, migration, India, reverse migration

[1] Assistant Professor, Jindal School of Government and Public Policy, O. P. Jindal Global University and founding member of the Centre for Complexity Economics, Applied Spirituality and Public Policy, O. P. Jindal Global University. Email: djha@jgu.edu.in

[2] Dean and Professor, School of Engineering and Technology, BML Munjal University; Founding member of the Centre for Complexity Economics, Applied Spirituality and Public Policy, O. P. Jindal Global University and Professor, School of Computational & Integrative Sciences, Jawaharlal Nehru University (on leave). Email: anirban.chakraborti@bmu.edu.in & achakraborti@jgu.edu.in

1. Introduction:

The Covid-19 pandemic has no doubt had serious consequences. Not only are the changes visible in the economy and the consumer markets, but the repercussions have also caused unemployment, hunger, and impoverishment at an unimaginable scale [Abi-Habib and Yasir, 2020]. This has affected the air quality, where some indicators like PM2.5 have reduced to as low as 40 percent [Neeraj Gupta, 2020]. The world however has been changing drastically since the pandemic struck, businesses, education, etc. have moved online. The lockdown was enforced soon after the pandemic struck in the hopes of limiting interaction and policies were framed to limit the negative externalities resulting from these interactions [Bethune and Korinek, 2020]. However, the closure of various sectors and restrictions in others was a big problem for developing countries such as India. Soon after the lockdown, we witnessed an exodus of migrants heading back to their native places both internationally and domestically [World Bank, 2020; Mitra et al., 2020]. Even the second wave had a repeat of the 2020 exodus but to a lesser extent. [Arabinda K. Padhee, Basanta K. Kar and Pranab R. Choudhury, 2020]

During the first wave, with a rising case count and death toll, the government started restricting international travel. The government made prior permission mandatory to cross to a different state amid the Covid-19 lockdown. This was done to curb the spread of the virus. However, an obvious negative externality of these policies was a lack of mobility for the many stranded workers in large cities across the country [Irudaya Rajan, S., Sivakumar, P. & Srinivasan, A., 2020]. This caused a financial and moral crisis, the result being the shramik train, special busses etc. that were made available to escort migrant workers to their native places [Telakat, 2020].

As noted above, the reverse migration resulting from Covid-19 and the lockdown caused a financial crisis. The contraction of the economy and the many closed businesses were not a solitary phenomenon limited to industries and the state, but it had a profound impact on the families of workers and their communities. This was especially true for low-income migrants whose families sustained themselves on remittances, often as the sole source of income [Behara, Mishra, 2021]. As most of these low-income migrants came from economically backward states, this leads to further socio-economic troubles. Furthermore, low-income migrants and their families in India often struggle for basic amenities; it is quite evident that they would not have been able to sustain themselves for long without either governmental assistance or employment opportunities. The onus once again comes to employment which was the primary reason the migrants left in the first place. Given that a substantial portion of the migrants was from rural areas [Census 2011] it could mean strain on the rural economy.

There are also places like tourist destinations, education hubs, and in general places that depended on the mobility of people for sustenance. A unique example of this would be the town of Kota in Rajasthan. The city runs on lakhs of students temporarily moving in for their education. As they leave, the prime source of income is lost [Chandra, 2020].

Another specific group likely to be severely affected by the pandemic and the ensuing lockdown is workers from the unorganized sectors. It is known that the number of jobs in the unorganized sector quantitatively far surpassed that of the organized sector. About 40 percent of India’s export is from the unorganized sector only. It is quite easy to apprehend that the effect of the pandemic is expected to be akin to the way financial crises had affected people in the unorganized sector [Estrada, 2020]. The 2009 global financial crisis had badly affected the unorganized sector in India; about 22 million jobs were lost in this sector since the onset of the crisis. Around this time, 100,000 Gem and jewelry-related jobs were lost in Gujrat, and 500,000 garment-related jobs were lost around Ludhiana and Tirupur [UNDP; Kumar et al. 2009].

A defining trait of low-earning migrants is a lack of social security, owing to the overflowing supply of their labour. This means that amid the recession these groups were especially vulnerable to a substantial decrease in quality of life. In parallel to the former issues, the same group has sustenance problems due to being able to earn barely more than a sustenance wage. This does not mean that all the migrants are bound to face similar problems; however, a substantial number of such migrants, especially those employed in the unorganized sector are expected to be impacted the most. Furthermore, workers especially migrants without written contracts and/or unskilled workers are even more likely to fall under such a category. A study on the migrant workers in Kerala amid the 2009 recession showed that more than 1/5th of the emigrants had to return to their state of origin while about 11.3 percent of them lost their livelihoods to the termination of their contract. Adding to the toll, another 3 percent had been compulsorily repatriated.

An important aspect of migration is its complexity and dependence on multiple factors often interacting nonlinearly. While relocation of migrants is deemed to be necessary, possibilities of circular migration can exacerbate the health crisis [Ginsburg et al., 2018]. Moreover, while the restrictions on mobility may not help impede the progression of the pandemic, such policies may end up imposing a negative externality [Carter, 2016]. In the context of the earlier SARS pandemic in 2009, it has been identified that the impact of lockdown and quarantine-related measures may impose a utility cost on the remaining population, thus acting as an incentive to vacate the quarantined locale [Mesnard & Seabright, 2009]. A similar argument with added baggage can be applied to the low-income migrants.

After the mass exodus amid deplorable conditions that had spurred the collective consciousness of the Indian public, the government responded with a draft on national migration labour policy [Mehrotra, 2021a]. This draft was formulated by NITI Aayog and the working subgroup of officials and members of civic society. The draft builds on the 2017 report’s views that call for comprehensive laws regarding the protection of the rights of workers in the unorganized sector which comprise a sizable portion of the migrant workforce, The draft marks out two viewpoints on policy design for the migrants. The initial approach consists of cash transfers and special quotas. While the second and the preferred method "enhances the agency and capability of the community and thereby removes aspects that come in the way of an individual's own natural ability to thrive". Some proposals include the setup of coordination mechanisms between the origin and destination states of migrants [Mehrotra, 2021b]. The draft also emphasizes on the loss of political representation for the migrants due to being registered in their state of origin. Hence some proposals include mechanisms to allow migrant participation, thus enabling political inclusion. Furthermore, use of Aadhar card for inclusion of migrants in social welfare programs which should remain accessible and undeterred across state borders [Niti Aayog, 2020]. A special unit under the ministry of labour and employment has also been suggested which would be responsible for the implementation of the aforementioned proposals and developing a map of focal points of migration and the set-up of migrant worker cells. A major difference from the 2017 report was the recommendation of raising the minimum wage in the origin states to "stem" migration.

In the present study, we have attempted to explore the relationship between pandemic-induced reverse migration of domestic migrants and Covid-19 cases in Indian states for both the first wave and the second wave of the Covid-19 pandemic. While it has been apparent from various media sources and policy discourses that the movement of the migrants from one state to another has been assumed to be a major source of the spread of the virus [Ray and Subramanian, 2020], academic study on the possible relationship between the reverse migration and the spread of the virus in Indian states is still limited, most probably due to the unavailability of an official estimate of the number of reverse migrants from different Indian states during this period. While there are three main sources of internal migration data in the country are – Census of India, NSSO surveys on migration, and the IHDS database, none of them provide an estimate of the number of migrants in different Indian states and districts in recent times, let alone the number of reverse migration induced by the recent pandemic. There are two main problems. First, while the principal source of migration data, the Census of India, provides numbers of migrants till 2010, the NSSO survey on migration was last conducted in the year 2007. Similarly, the most recent IHDS-II survey provides information on migrant laborers until 2011-12. Thus all these data sources are backdated. Second, often these three data databases provide different estimates of the number of internal migrants due to the differences in the methodology of the surveys. Hence, there are two objectives of this study. First, we want to construct a dataset that can provide a reliable estimate of pandemic-induced mobility of people from different parts of the country using train running information on the Covid-19 special trains. Second, we conduct an econometric exercise to uncover the relationship between the mobility of people from Maharashtra, one of the largest migrant-receiving states, to different Indian states during the second wave of the pandemic using train passenger data constructed in the first part. We have focused on the reverse migration from Maharashtra because the second wave of Covid-19 in Maharashtra has started at least two months ahead of the second wave in other Indian states [see Figure 8].

The rest of the paper has been organized as follows. In section 2 first, we discuss the list of different data used in this study along with their sources. In this section, we also discuss the method to construct the train passenger mobility data among the Indian states using the

Covid-19 special train information following a complex network approach. The complex network diagrams constructed from the Census-2011 data and train passenger mobility data have been presented in Section 3. We present the detailed specification of the econometric models and the results of the econometric exercise in Section 4. Section 5 concludes the study with a discussion on the possible public policies to tackle the complex issue like reverse migration induced health crisis in the country.

2. Data and Methodology

To understand the existing network of mobility of people before the pandemic we have used Census-2011 data from the census of India. The census uses a two-step process. It is in the second phase that the data regarding migration and its reasons were collected. The data is available as an adjacency matrix and was prepared for use with some data cleaning.

In order to construct the Covid-19 induced reverse migration data, we have used publicly available information on Covid-19 special trains from India Rail Info. Because the ministry of railways does not provide detailed information on the mobility of passengers, we have scraped the train running information of Covid-19 special trains from the website https://indiarailinfo.com/ using a web crawler. Then the scraped data has been used to construct a complex mobility network of train passengers in the entire country for the period 1st June 2020 to 7th June 2021. The aforementioned website is a crowdsourced database holding information regarding all types of trains, and currently hosts the data regarding trains run since the advent of the Covid-19 spurred lockdown. The database is being updated continuously.

The data for the confirmed Covid-19 cases have been collected from the website of Covid-19 India Org (https://www.Covid19india.org/). The website stores data for daily cases and total cases for recovered, deceased, confirmed, and active cases. Finally, the data on population density, the fraction of urban population, gross state domestic product per capita, etc have been taken from the Handbook of Statistics on Indian States, Reserve Bank of India, while the state-level data on health infrastructure and health personnel have been collected from Health Profile of India, Government of India.

2.1 Methodology of Network Construction

The nodes in the network represent states while the outgoing edge represents emigration and the incoming edge represents immigration. The total numbers of nodes are 35 including states and union territories. The number is fewer than the current number of states and union territories due to Telangana and Andhra Pradesh being treated like one. This was done due to a lack of explicit data for Telangana in Census-2011 data which we are using. The same can be said about Ladakh and Jammu and Kashmir which were under the same jurisdiction before August 2019.

We follow the algorithm used in the construction of a complex migration network for India in Onkar Sadekar, Mansi Budamagunta, G. J. Sreejith, Sachin Jain, M. S. Santhanam [2021]. In constructing the train passenger mobility network we used two assumptions. First, all the trains run at full capacity. Second, the number of people who move from state *i* to state *j* is proportional to their populations.

The resultant formulae are as follows

$$F_a^b = \left[C\,\frac{N_a}{\sum_{j=a}^{k} N_j}\right]\left[\frac{N_b}{\sum_{j=b+1}^{k} N_j}\right], \qquad \text{for,} \quad 2 \le a < b \le k,$$

$$F_a^b = \left[C\frac{N_b}{\sum_{j=a+1}^{k} N_j}\right], \qquad \text{for,} \quad 1 = a < b \le k.$$

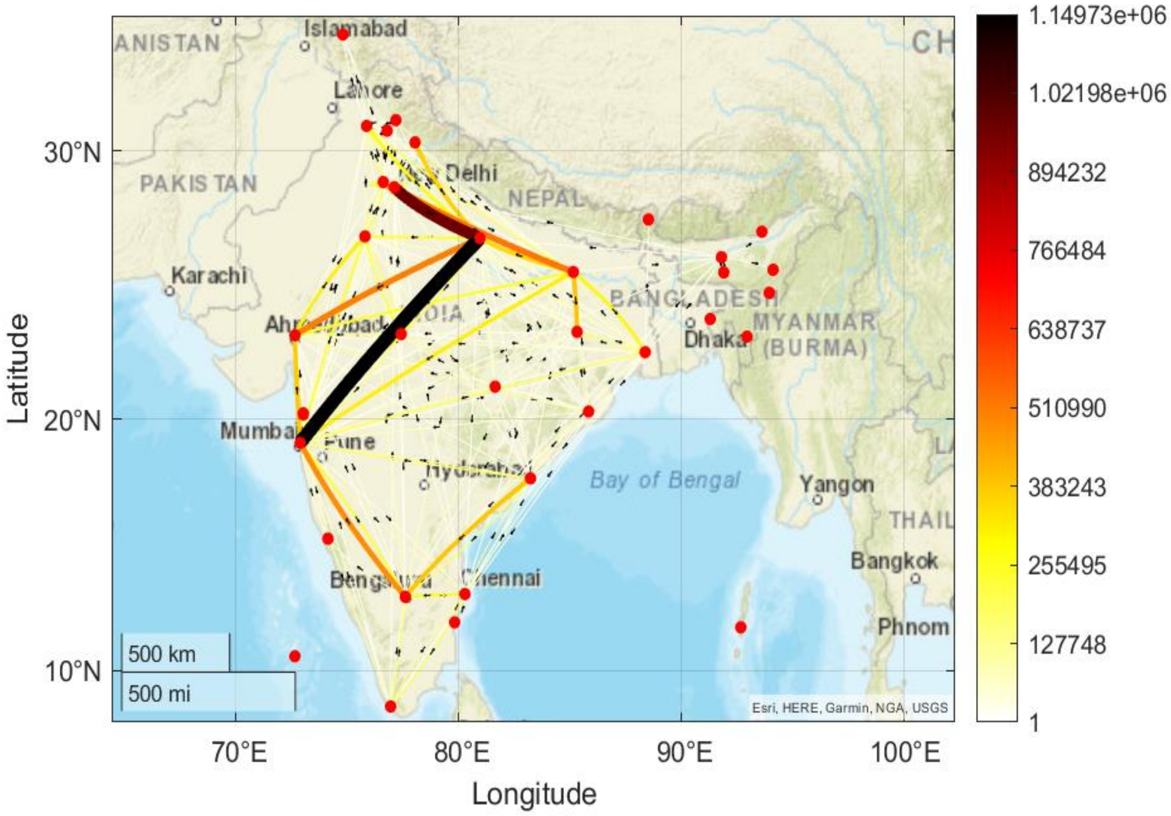
Here *F(a,b)* is the movement from the state *a* to state *b*. *C* is the total capacity of the trains. N_i is the population in state *i*. For getting the capacity of a train we needed the days it had run and the capacity it had per run that is the seating capacity of the bogies together. The seating capacity was calculated using the data for bogies which were available as the rake position. However, this data was not available for all trains hence we assigned the average seating capacity as the seating capacity for the trains where the rake position is missing. The individual bogie types and the number of their kind on the train being iterated on were used to decide the seating capacity. This data regarding the bogie type and capacity was taken from the website https://www.trainman.in/. After getting the total capacity using the days the train had run, we used the prior mentioned formulae to calculate the edges of the network. To calculate the days the train had run, we used the date of the first run and the date of the last run. This data was not available for all the trains. So to construct the network, we have only used those trains for which the date of the first and last run was available. This shortened the number of trains used in forming the network; hence we have used only 36 percent of a total of 2832 trains running during the entire period. Also note that when we have constructed the network for shorter periods, the capacity variable is altered according to the amount of overlapping running days of each train with the start and end dates we are considering.

3. Visualization of Complex Network of Migration

In order to understand the pattern of interstate migration of individuals between different Indian states before the pandemic, first, we present the network plots constructed using Census-2011 migration data in section 3.1. In Section 3.2 we present the network plots constructed using the data on Covid-19 special trains for two periods - 1st of June 2020 to 7th June 2021 [full sample] and 14th Feb 2021 to 14th May 2021 [Second Wave]. Finally, in Section 3.3 we discuss the nature of a possible lead-lag relationship between reverse migration from Maharashtra to other Indian states and Covid-19 cases in those states.

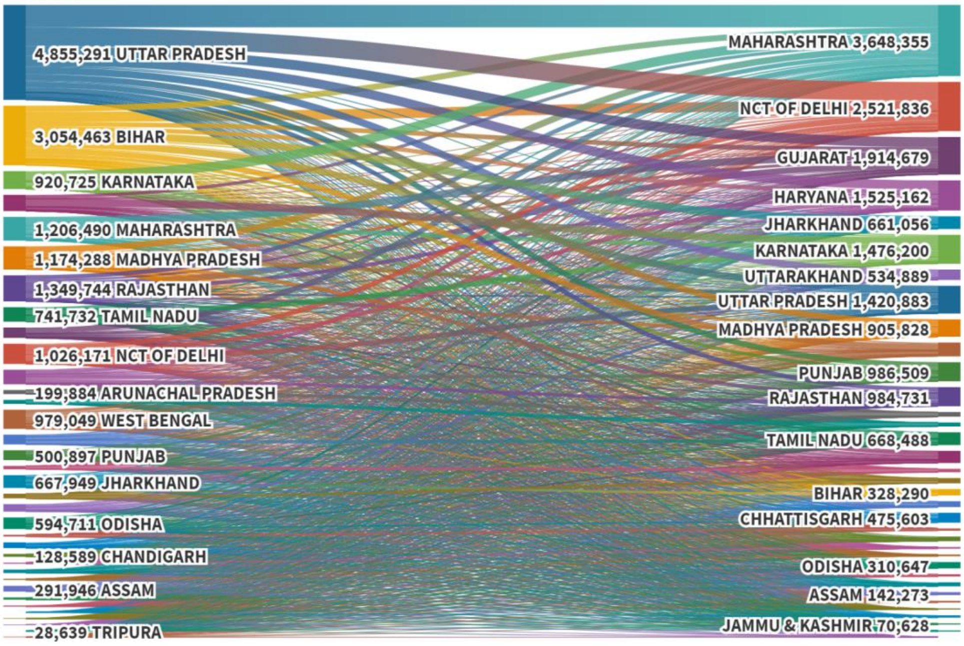
3.1 Network plots using the Census-2011 data

Figure 1: Interstate migration network before the pandemic using Census-2011 data



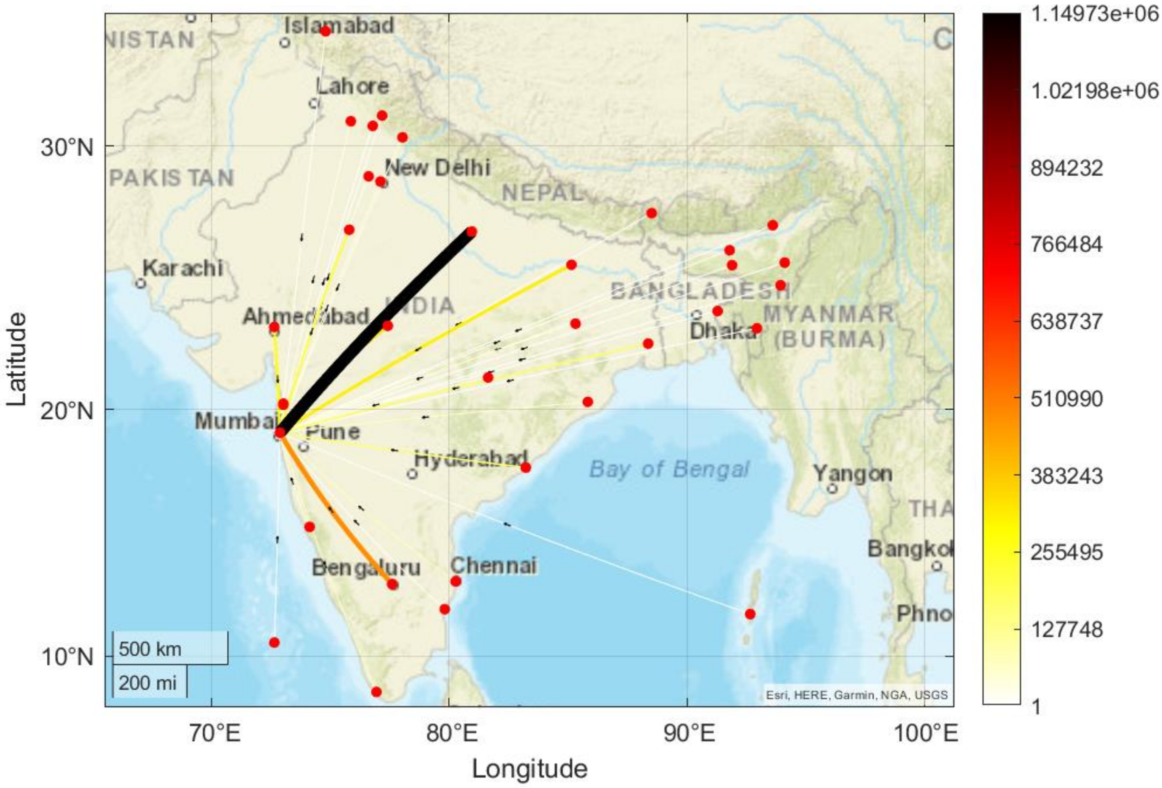
Source: Author's estimation

Figure 2: Sankey Diagram using Census-2011 data



Source: Author's estimation

Figure 3: In-migration in Maharashtra using Census-2011 data



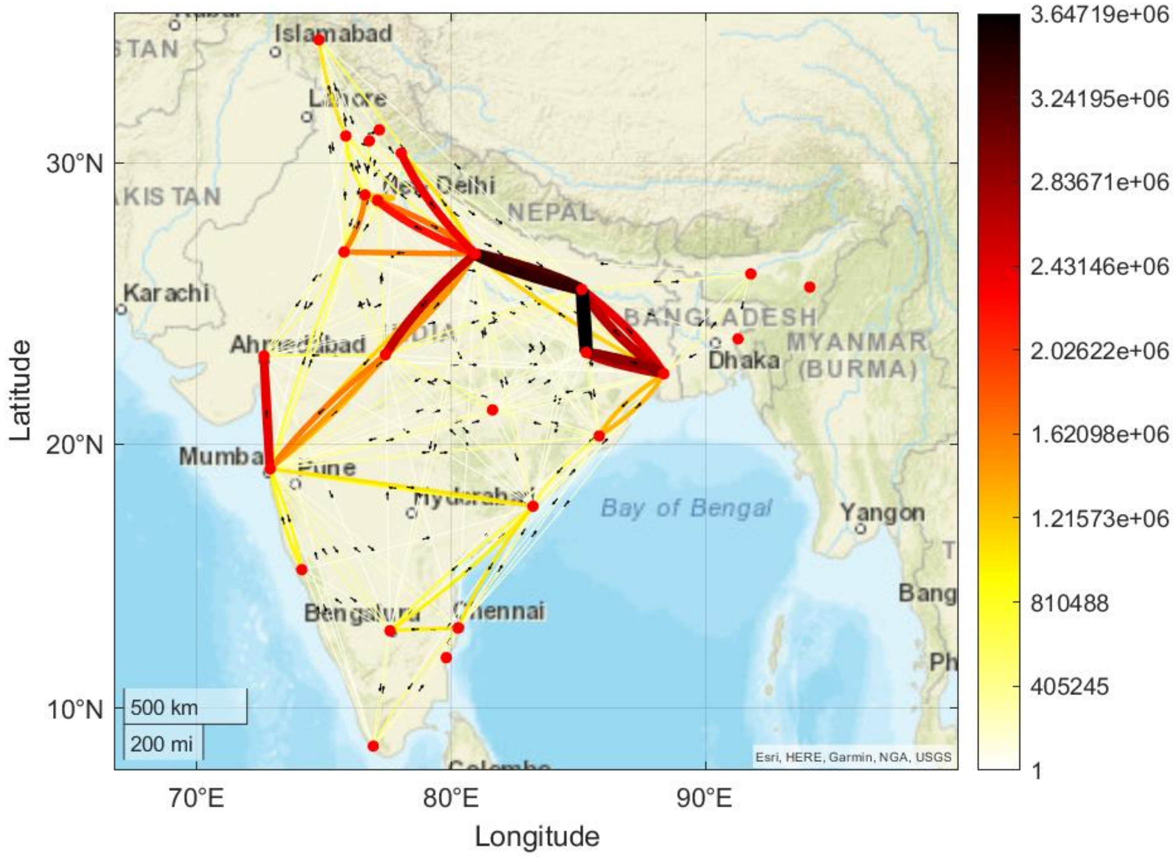
Source: Author's estimation

The network for the 2011 census migration is given above. For the sake of visualizing better, we have only included the first 200 edges in this plot as in the next. The most prominent edges are the connections from Uttar Pradesh to Maharashtra and from Uttar Pradesh to Delhi. These are shown in brown and black colours respectively and have around 1 million migrants each. This is not surprising as Uttar Pradesh has the highest quantity of intra-national emigrants [Census of India 2011]. Following these, we have connections from Bihar to Delhi, Uttar Pradesh to Gujarat, and Karnataka to Maharashtra. As is apparent from the network, Maharashtra has the most immigrants. There are 8 states from where Maharashtra has more than 100 thousand immigrants. In the network, this has been represented by edges with yellow tints and darker tints. A similar number is seen for Uttar Pradesh but the flow was opposite. Uttar Pradesh has 8 states for which it has more than 100 thousand emigrants, represented in the same manner in the network with arrows reversed. However, unlike Bihar, Uttar Pradesh has substantial in-migration itself. For the 2011 census, Uttar Pradesh, Bihar, Rajasthan, and Madhya Pradesh [in order of quantity of emigrants] had more than 50 percent emigrants. While states like Maharashtra, Delhi, Gujarat, Uttar Pradesh, and Haryana account for 50 percent of the total immigrants. Note that Uttar Pradesh is a part of both lists. This has been further demonstrated in the Sankey diagrams plotted for the same data. Sankey diagrams are used to represent flow and the width of the lines represent the flow rate used, mainly used in engineering. Here the flow rate is synonymous with migrants moving from the node on the left side [out-migrating state] to the node on the right [in-migrating state]. Along with the complete census network, we have included a Maharashtra-specific network, which shows the states from where migration occurs in Maharashtra.

These trends have been consistent over the decades and several studies have been conducted to understand such behavior with the lens of various competing models. One of the most popular models, Lee's model has been originally introduced in [Lee, 1966]. [Rani, S. 2018; Wang 2010; Massey 1993] are some other works in this pursuit. In India, migration has been on the rise since 1981 [Raja, Bhagat, 2021]. The list of states with high out-migration and in-migration of the population has not shown much change [Das, Saha 2021].

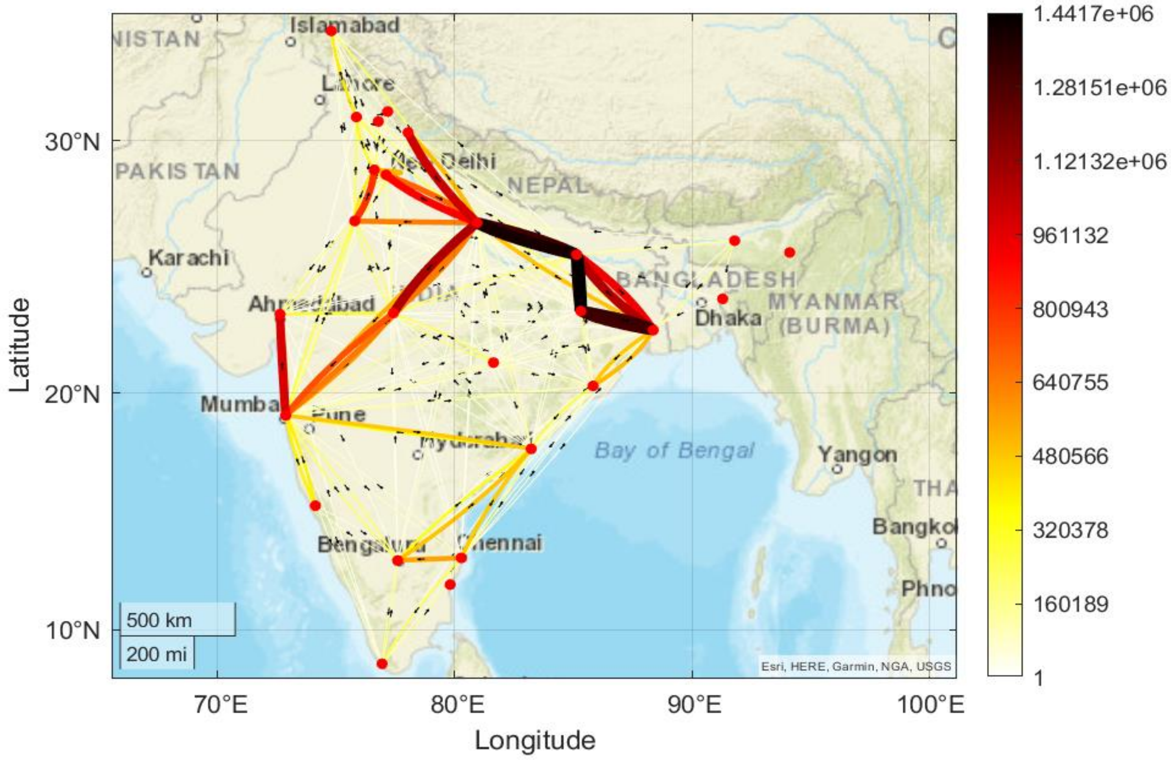
3.2 Network plots using Covid-19 special train data

Figure 4: Full sample network using Covid-19 special train data



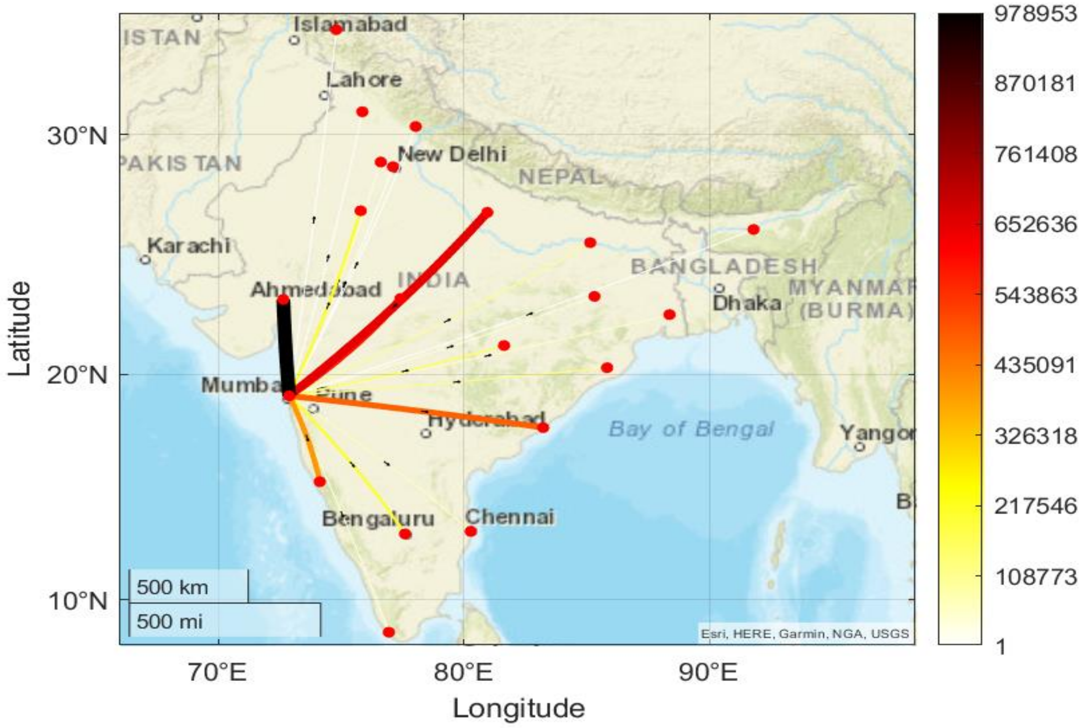
Source: Author's estimation

Figure 5: Train mobility network for data between 14th Feb and 28th June



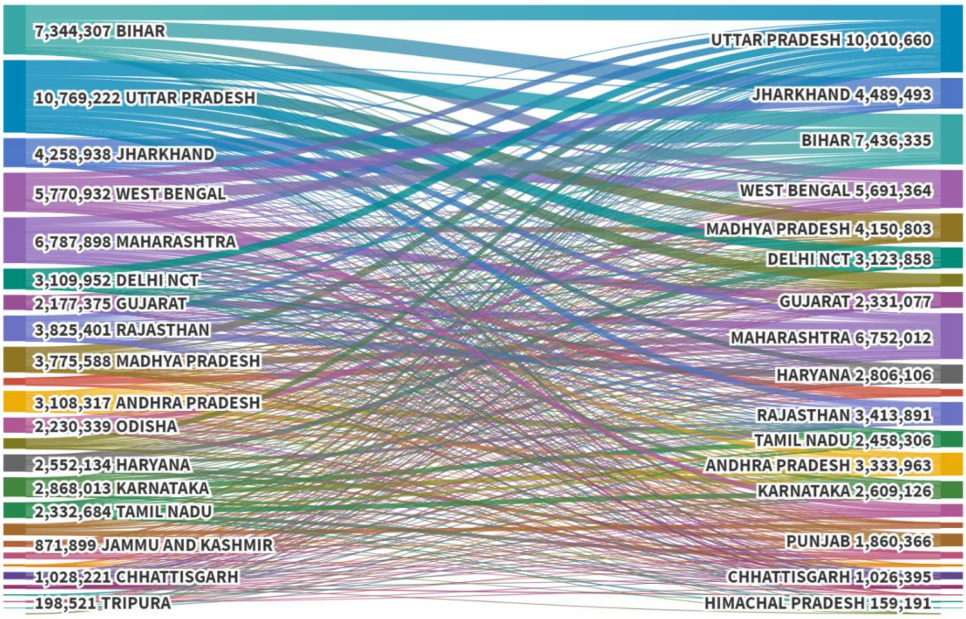
Source: Author's estimation

Figure 6: Network for outmigration from Maharashtra between 14th Feb and 14th May



Source: Author's estimation

Figure 7: Sankey diagram for the train dataset



Source: Author's estimation

The network constructed from the Covid-19 special trains is then plotted for two time periods. The first one is for the entire period and the other one is only for the second wave. Just like for the Census-2011 data, the entire network is very dense, so only the top 200 edges based on the weights are shown. Each edge is between two states and the nodes corresponding to each state are marked in its capital city. The width of the line as well as the color of each edge is drawn according to the weight of that edge. Thicker and darker (Towards the black) edges represent the connections with higher weights and the thinner and lighter (Towards the yellowish-white) edges represent the ones with the lower weights. Both the networks show heavy connections between the states of Maharashtra, Gujarat, Bihar, Uttar Pradesh, and Delhi. In the case of the second wave we have chosen only those trains that have run between the 14th February and 14th May then the same procedure of visualization is followed. Sankey plots have been used as well to demonstrate the total movement of people between the states. Similar to the Census-2011 data, the width for a state represents the total mobility of people from that state and the right side shows the total mobility to that state.

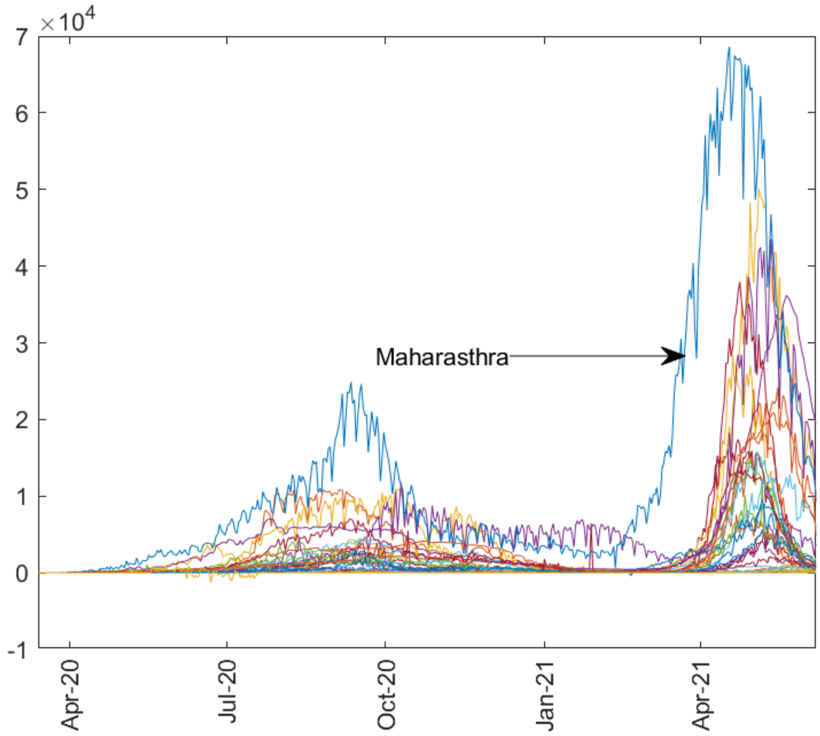
Since we are interested in the role of Maharashtra particularly, the outgoing links from Maharashtra alone are plotted. There were 20 non-zero edges.

3.3 Possible relationship between Covid-19 cases and migrants

In Figure 8 the time series plots have shown a clear lead-lag relationship in the Covid confirmed data for individual states, where many of the states are observed lagging behind Maharashtra. Maharashtra is also observed to be a major node in the migration network with very strong links with other important states such as Uttar Pradesh and Delhi. In order to explore this relationship between migration and Covid-19, this figure plots the number of people who have travelled to different states from Maharashtra and the total number of confirmed cases in those states in the two axes. We are using the proxy data constructed from the COVID special express train information as a measure of migration between the states. And the daily confirmed cases data is integrated for the required period to calculate the total number of Covid-19 cases reported. Two scatter plots [Figure 10 and 11] are made for the two time periods that we are interested in which are a) 14th Feb to 14th May [From beginning to the end of the second wave] and b) 14th Feb to 28th Jun [From the beginning of the second-wave to the last available date].

Along with the data points, a linear fit is also shown to observe any general trend and the correlation coefficient between the two variables is calculated to see the strength of the relationship.

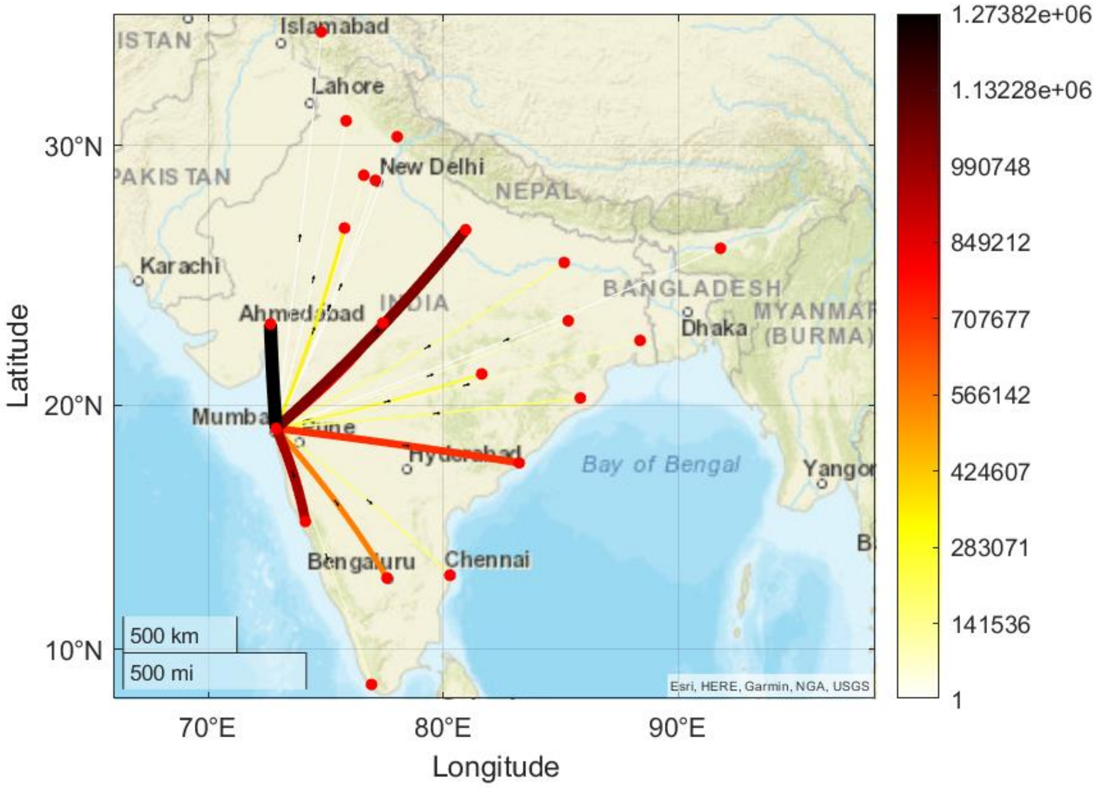
Figure 8: Time plot of Covid-19 cases in Indian states



Source: Author's estimation

Figure 8 shows confirmed cases for different states in India. It is a time series plot showing the various phases of the Covid-19 pandemic that unfolded in India for all the different states. Since Maharashtra is the state with the highest number of immigrants and at the same time, the highest contributor state to the number of total Covid-19 cases in India, we will focus on that more staunchly. The start of the first wave in India amid lockdown was visibly slower than the second, and we can see a simultaneous increase in the numbers for every state. The peak of the first wave was reached around 11th September 2020 which was also the peak of the first wave in Maharashtra. This changed for the second wave when Maharashtra bore witness to its second wave much earlier than the rest of India. While the case count was stable for most of the states and even practically disappeared for some states like Bihar which reported fewer than 100 cases for weeks. Nearly a month after cases started rising in Maharashtra, other states started reporting a rise in confirmed cases count as well; effect seen in the constructed network too (vide Figure 9). Several vouched for caution and depicted the possibility of a second wave [Yasmeen 2021]. At the same time, these concerns were unwarranted to some even after a surge in cases in Maharashtra [Noronha, 2021]. It is also important to note that because Delhi already had 2 such upswings, it is understandable where such views come from.

Figure 9: Migration from Maharashtra using Covid-19 special train data



Source: Author's estimation

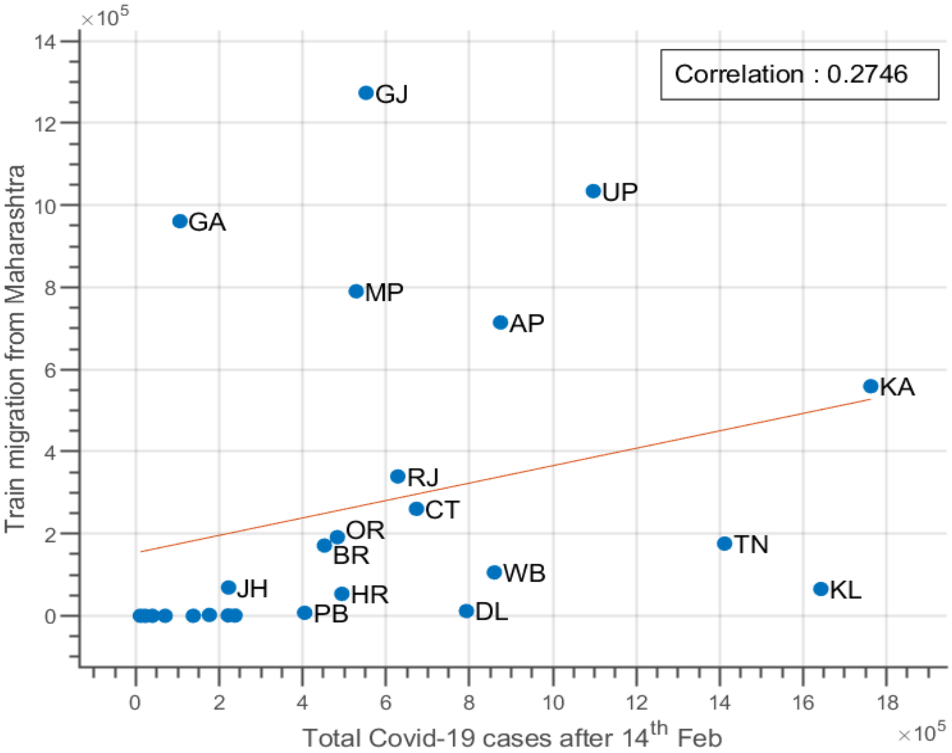
4. Relationship between migration and spread to Covid-19

In order to find out the relationship between reverse migration and the spread of Covid-19 cases in Indian states during the second wave of Covid-19 spread, we relied on the ordinary least square (OLS) regression approach. In this study, we have taken only the 20 Indian states which are connected through a train network. The dependent variable in all the models is the total number of confirmed Covid-19 cases per lakh population in different Indian states between 14th Feb 2021 and 14th May 2021. The independent variables used in the model are –

i. The number of train passengers travelled from Maharashtra per lakh population of a state during this period.
ii. Number of migrants from different Indian states to Maharashtra.
iii. Population density.
iv. Percentage of the urban population.
v. Hospital beds per lakh population.
vi. Doctors per lakh population.
vii. Log gross state domestic product per capita

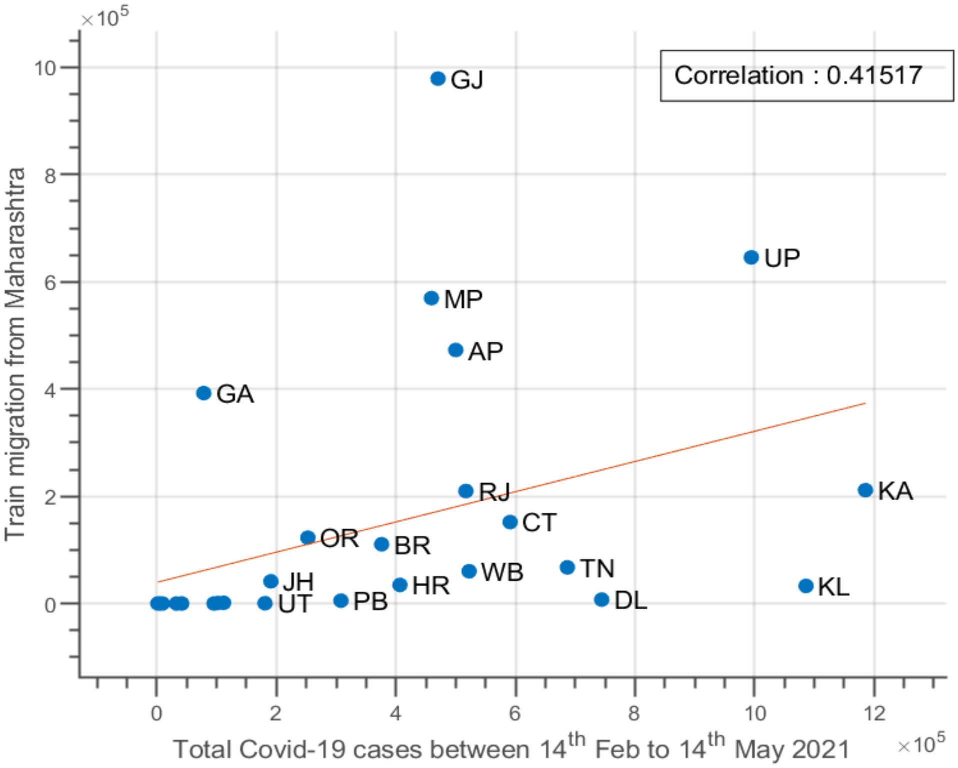
The first two independent variables try to capture the reverse migration of labourers from Maharashtra to different Indian states. The hypothesis for taking these two variables is that the virus can spread from one state to another only through human transmission. Since migrant workers return to their place of origin whenever a lockdown has been imposed, these returning migrants may be one of the crucial factors behind the spread of the virus from one state to another besides international migrants. Studies have already shown a positive and significant relationship between international migrants and the spread of the virus in the district of India during the first 45-days of the onset of the pandemic. The first variable is the number of train passengers travelling from Maharashtra to different Indian states per lakh population of that state. Since the train is one of the most important mediums of transport for the migrant workers in the country [Economic Survey, 2017] and the second wave of the pandemic has started in Maharashtra almost two months ahead of other Indian states [see Figure 8], reverse migration from Maharashtra may be one of the principal sources of the spread of the virus from Maharashtra to other Indian states. Accordingly, we expect a positive impact of this variable on the spread of Covid-19 cases in Indian states. The second variable is the number of 0-9 year migrants in Maharashtra from different Indian states per lakh population of that state. Since the exodus of migrant workers has been perceived to be one of the crucial spreaders of the virus to different states and Maharashtra is one of the largest migrant-receiving states in the country, the reverse migration from Maharashtra is expected to be high. Accordingly, we can expect a positive relationship between the spread of Covid-19 cases in different Indian states and the number of migrants in Maharashtra from other states. For an initial assessment of the possible relationship between these two variables, we present two scatter plots of the number of train passengers travelled to different Indian states from Maharashtra during the second wave of the pandemic and Covid-19 confirmed cases during the same period in Figure 10 and 11. Both the figures show a positive relationship between confirmed cases and train passenger mobility data during the second wave of the pandemic.

Figure 10: Reverse migration from Maharashtra vs. Covid-19 confirmed cases from 14th Feb 2021 to the end of the sample



Source: Author's estimation

Figure 11: Reverse migration from Maharashtra vs. Covid-19 confirmed cases between 14th Feb 2021 and 14th May 2021



Source: Author's estimation

In different studies on the spread of the infectious virus during earlier epidemics, it has been found that population density helps the virus to spread from one person to another. Similarly, the population density is more in urban areas relative to rural areas. Hence, population density per square kilometre and percentage of urbanization population in a state may also be important determinants of the Covid-19 virus spread. Therefore, we can expect a positive impact of these two variables on the total Covid-19 cases per lakh population in the regression models.

Moreover, the availability of health infrastructures such as the number of hospital beds and the availability of health personnel such as doctors, nurses can have a negative impact on the spread of the virus. In order to treat the infected persons and quarantine suspected Covid-19 patients, we need more hospital beds. Hence the availability of hospital beds may reduce the spread of the virus. In the states where more hospital beds are available per lakh population the chance of spread of the virus is less. Similarly, the availability of doctors can help in the dissemination of crucial information regarding the safety norms during a pandemic. Hence, these two variables, hospital beds per lakh population and availability of doctors per lakh population, are expected to have a negative relationship with Covid-19 cases in a state.

The last independent variable that we have considered in this study is log per capita gross state domestic product. Higher per capita income states can spend a higher amount of state resources to curb the spread of the virus. Accordingly, we can expect a negative association between log per capita income and Covid-19 cases. However, more urbanized and densely populated states are found to have higher per capita income in this country. Hence, higher per capita income and Covid-19 cases can also have a positive relationship with Covid-19 cases in a state. Hence, it is difficult to predict the expected sign of this variable.

Based on the above discussion the regression model that we have used in this study is as follows –

$$Y = f(TMIG, CMIG, URPOP, POPDEN, INFBED, INFDOC, PCI) \tag{1}$$

where *Y* is our dependent variable which denotes Covid-19 cases per lakh population in the states. *TMIG* represents the number of train passengers per lakh population traveled from Maharashtra to another Indian state during the second wave of the pandemic. *CMIG* is the number of 0-9 year migrants in Maharashtra from other Indian states per lakh population of the respective state. *URPOP* represents the percentage of the urban population in a state. Similarly, *POPDEN* denotes the density of population in a state. *INFBED* is a health infrastructure variable and represents the number of hospital beds available per lakh population in a state. *INFDOC* represents the number of doctors available per lakh population. *PCI* denotes log per capita GSDP in a state. Based on the above equation we have carried out two sets of regression models. In the first set, the main independent variable is *TMIG*, i.e., the number of train passengers per lakh population traveled from Maharashtra to another Indian state during the second wave of the pandemic. In the second set, the main independent variable is *CMIG* i.e., the number of 0-9 year migrants in Maharashtra from other Indian states per lakh population of the respective state. In both the first and second set of regression models, we have carried out six different regression models. Therefore, the functional forms of the first set of our regression models are as follows –

$$Y = \beta_0 + \beta_1 TMIG + \beta_2 URPOP + \beta_3 PCI + \varepsilon_1 \quad (2)$$

$$Y = \beta_0 + \beta_1 TMIG + \beta_2 URPOP + \beta_3 PCI + \beta_4 INFBED + \varepsilon_2 \quad (3)$$

$$Y = \beta_0 + \beta_1 TMIG + \beta_2 URPOP + \beta_3 PCI + \beta_4 INFBED + \beta_5 INFDOC + \varepsilon_3 \quad (4)$$

$$Y = \beta_0 + \beta_1 TMIG + \beta_2 POPDEN + \beta_3 PCI + \varepsilon_4 \quad (5)$$

$$Y = \beta_0 + \beta_1 TMIG + \beta_2 POPDEN + \beta_3 PCI + \beta_4 INFBED + \varepsilon_5 \quad (6)$$

$$Y = \beta_0 + \beta_1 TMIG + \beta_2 POPDEN + \beta_3 PCI + \beta_4 INFBED + \beta_5 INFDOC + \varepsilon_6 \quad (7)$$

Table 1: Regression results using train migration data

Var.	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
TMIG	0.13 (0.022)*	0.12 (0.022)*	0.13 (0.023)*	0.14 (0.029)*	0.14 (0.030)*	0.15 (0.030)*
URPOP	34.06 (6.743)*	33.72 (6.950)*	33.86 (8.621)*			
POPDEN				0.16 (0.054)*	0.15 (0.058)**	0.16 (0.056)**
PCI	1844.77 (966.57)***	1916.70 (973.93)***	1876.04 (1078.589)***	3257.97 (1047.556)*	3304.58 (1074.51)**	3123.08 (966.629)**
INFBED		-1.25 (1.723)	- 1.29 (1.699)		-0.67 (1.69)	- 0.85 (1.721)
INFDOC			0.18 (3.695)			0.89 (3.708)
Cons.	-8847.08 (4694.193)***	-9095.68 (4748.61)***	-9466.29 (1034.730)***	-15096.79 (5117.604)*	-15272.98 (5259.95)*	-14424.15 (4657.31)*
R-squared	0.7662	0.7692	0.7692	0.7601	0.7609	0.7614

*Note: *, **, and *** represents significance at 1%, 5% and 10% level respectively*

The results of the above regression models have been presented in columns 2-7 respectively in Table 1. Since both urban population and population density have a similar impact and may have a high correlation, we have not taken them in the same regression model. While *URPOP* has been taken in models 1-3, in models 4-6 we have used *POPDEN*. In all the models the main independent variable TMIG is found to be positive and significant. This suggests that an increase in one more person traveled from Maharashtra through train per lakh population in a state would increase 0.13 number of Covid-19 cases per lakh population of that state in model 1. While *POPDEN, URPOP,* and *PCI* have been found to be positive and significant, *INFBED* and *INFDOC* are insignificant. This indicates that densely populated, more urban, and richer states have a higher chance of spread of the virus. Since most of the high per capita income states in India are more urbanized and have relatively more population density, the positive sign of *PCI* is not surprising. The sign of *INFBED* is as expected negative. This indicates that the availability of more hospital beds can help to reduce the spread of the virus. While *INFDOC* is insignificant, the sign is positive. This may be due to the fact that the availability of more doctors per lakh population can enhance the detection of Covid-19 cases.

As discussed above, we have used an alternate variable to represent the reverse migration of the labourers from Maharashtra using Census-2011 data in the second set of regression models viz. *CMIG*. The functional forms of these regression models are as follows –

$$Y = \beta_0 + \beta_1 CMIG + \beta_2 URPOP + \beta_3 PCI + \varepsilon_7 \tag{8}$$

$$Y = \beta_0 + \beta_1 CMIG + \beta_2 URPOP + \beta_3 PCI + \beta_4 INFBED + \varepsilon_8 \tag{9}$$

$$Y = \beta_0 + \beta_1 CMIG + \beta_2 URPOP + \beta_3 PCI + \beta_4 INFBED + \beta_5 INFDOC + \varepsilon_9 \tag{10}$$

$$Y = \beta_0 + \beta_1 CMIG + \beta_2 POPDEN + \beta_3 PCI + \varepsilon_{10} \tag{11}$$

$$Y = \beta_0 + \beta_1 CMIG + \beta_2 POPDEN + \beta_3 PCI + \beta_4 INFBED + \varepsilon_{11} \tag{12}$$

$$Y = \beta_0 + \beta_1 CMIG + \beta_2 POPDEN + \beta_3 PCI + \beta_4 INFBED + \beta_5 INFDOC + \varepsilon_{12} \tag{13}$$

The results of these regression models have been presented in Table 2. In these models as well the independent variables are the same as in the first set of regression models. Similar to the first set, we have used *URPOP* in the first three regression models and *POPDEN* in the final three models due to the possibility of a high correlation between these two variables.

Table 2: Regression results using Census-2011 migration data

Var.	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
CMIG	1.90 (0.832)**	2.17 (0.603)**	2.40 (0.547)*	2.07 (0.878)*	2.32 (0.673)*	2.62 (0.520)*
URPOP	29.43 (7.961)*	28.99 (7.56)*	29.94 (8.320)**			
POPDEN				0.11 (0.052)**	0.11 (0.51)**	0.12 (0.044)**
PCI	2721.23 (1044.891)**	2638.626 (850.038)*	2308.256 (793.747)**	1033.70 (3.92)*	3985.91 (970.63)*	3603.76 (688.249)*
INFBED		-4.78 (1.379)*	-5.51 (1.636)*		-4.66 (1.449)*	-5.56 (1.738)*
INFDOC			1.68 (3.211)			2.11 (2.963)
Cons.	-13581.79 (5110.27)**	-12841.97 (4116.961)*	-11347. 15 (3795.941)*	-19569.33 (5074.30)*	-18916.43 (4694.912)*	-17181.97 (3425.08)*
R-squared	0.7097	0.7548	0.7563	0.6985	0.7410	0.7433

*Note: *, **, and *** represents significance at 1%, 5% and 10% level respectively*

The results of the second set of regression models are very similar to the first set of the models. In all the regression models, the main variable of concern (*CMIG*) has been found to be positive and significant. In model 7, this indicates that an increase in one more migrant per lakh population from an Indian state to Maharashtra is expected to increase 1.90 more number of Covid-19 cases in an average Indian state. Similarly, while the coefficient of *URPOP*, *POPDEN*, and *PCI* are found to be positive and significant, *INFDOC* is also positive but insignificant. The only difference we found in the second set of regression models from the first set of models is the coefficient of *INFBED*. Here, the coefficient is negative and significant in all the models. This suggests that an increase in the number of hospital beds per lakh population may reduce the spread of the virus.

5. Conclusion and policy recommendations

One of the major problems that India faced in recent times, which emerged after the imposition of initial lockdown in March 2020 to stop the spread of the Covid-19 virus, is the reverse migration of labourers and the associated anxiety of the spread of the virus in the origin states of the migrants from the return migrants. The issue of reverse migration is so complex that most of the policies undertaken by the government have failed to produce any fruitful outcome. While it has been perceived that reverse migration of the labourers is associated with the spread of the virus, academic research to confirm this hypothesis is still limited. We identified that one of the possible reasons may be the unavailability of data on the number of reverse migrants in different states. Accordingly, in this study, we tried to construct a dataset on reverse migration of the interstate migrants in India from the information available on the Covid-19 special trains using a complex network approach. Then we applied an econometric exercise to understand the possible relationship between the confirmed Covid-19 cases in different Indian states and the reverse migration data that we have constructed during the second wave of the pandemic. In order to check the robustness of the results we also carried out a similar econometric exercise using the Census-2011 migration data.

The results derived from the econometric models identified a positive and significant relationship between the reverse migration from Maharashtra and the Covid-19 cases in different Indian states during the second wave of the pandemic. Moreover, we found that while population density and the fraction of urban population are also positively related to the Covid-19 cases in the states, the availability of hospital beds has a negative relationship with the number of Covid-19 cases in the Indian states. Therefore, to tackle the crisis and the spread of the virus government should focus on the problems of reverse migrants and increasing the health infrastructures like hospital beds, isolation centres, availability of ventilators, etc.

Acknowledgments

We acknowledge the contributions of Hrishidev and Suhaas Neel in preparing the data and network plots for this draft.

Reference:

1. 2011 Census data: (2021). Retrieved 5 July 2021, from https://censusindia.gov.in/2011census/migration.html/
2. Abi-Habib, M., & Yasir, S. (2021). India's Coronavirus Lockdown Leaves Vast Numbers Stranded and Hungry (Published 2020). Retrieved 5 July 2021, from https://www.nytimes.com/2020/03/29/world/asia/coronavirus-india-migrants.html
3. Gupta, N., Tomar, A., & Kumar, V. (2021). The effect of COVID-19 lockdown on the air environment in India. Retrieved 5 July 2021, from http://dx.doi.org/10.22034/GJESM.2019.06.SI.04
4. Bethune, Z. A., & Korinek, A. (2020). Covid-19 infection externalities: Trading off lives vs. livelihoods. Technical report, National Bureau of Economic Research.
5. Padhee, Arabinda K., Kar, Basanta K., & Choudhury, Pranab R. (2021). The Lockdown Revealed the Extent of Poverty and Misery Faced by Migrant Workers. Retrieved 5 July 2021, from https://thewire.in/labour/Covid-19-poverty-migrant-workers
6. Ray, D., & Subramanian, S. (2020). India's lockdown: An interim report. Indian Economic Review, 55(1), 31-79.

7. Rajan, S. I., Sivakumar, P., & Srinivasan, A. (2020). The COVID-19 pandemic and internal labour migration in India: A ‘crisis of mobility’. The Indian Journal of Labour Economics, 63(4), 1021-1039.
8. 1,500 buses deployed to take migrant workers to their native districts: Ghaziabad DM. Retrieved 5 July 2021, from https://economictimes.indiatimes.com/news/politics-and-nation/1500-buses-deployed-to-take-migrant-workers-to-their-native-districts-ghaziabad-dm/articleshow/74874985.cms
9. Behera, M., Mishra, S., & Behera, A. R. (2021). The COVID-19-Led Reverse Migration on Labour Supply in Rural Economy: Challenges, Opportunities and Road Ahead in Odisha. The Indian Economic Journal, 00194662211013216.
10. COVID-19 lockdown leaves Kota's coaching centres deserted; shift to online classes offers only hope-India News , Firstpost. (2021). Retrieved 5 July 2021, from https://www.firstpost.com/india/Covid-19-lockdown-leaves-kotas-coaching-centres-deserted-shift-to-online-classes-offers-only-hope-8664341.html
11. Estrada, M. A. R. (2020). The Difference between The Worldwide Pandemic Economic Crisis (COVID-19) And The Global Financial Crisis (Year 2008).
12. Ginsburg, C., Collinson, M. A., Gómez-Olivé, F. X., Gross, M., Harawa, S., Lurie, M. N., ... & White, M. J. (2021). Internal migration and health in South Africa: determinants of healthcare utilisation in a young adult cohort. BMC public health, 21(1), 1-15.
13. Kumar, R., Debroy, B., Ghosh, J., Mahajan, V., & Prabhu, K. S. (2009). Global Financial Crisis: Impact on India's Poor Some Initial Perspectives. Retrieved October, 15, 2018.
14. Carter, D. (2016). Hustle and brand: The sociotechnical shaping of influence. Social Media Society, 2(3), 2056305116666305.
15. Mesnard, A., & Seabright, P. (2009). Escaping epidemics through migration? Quarantine measures under incomplete information about infection risk. Journal of Public Economics, 93(7-8), 931-938.
16. Mehrotra K. (2021). Explained: What is NITI Aayog’s draft national policy on migrant workers?. Retrieved 5 July 2021, from https://indianexpress.com/article/explained/niti-aayog-migrant-workers-policy-Covid-lockdown-7201753/
17. Mehrotra K. (2021). Enable vote, map skill, inter-state teams: First migrant policy draft.. Retrieved 5 July 2021, from https://indianexpress.com/article/india/migrant-worker-first-draft-policy-niti-aayog-labour-ministry-Covid-19-7197599/
18. Sadekar, O., Budamagunta, M., Sreejith, G. J., Jain, S., & Santhanam, M. S. (2021). An infectious diseases hazard map for India based on mobility and transportation networks. arXiv preprint arXiv:2105.15123.
19. Rajan, S., Irudaya., & Bhagat, R., B., (2021). Internal Migration in India: Integrating Migration with Development and Urbanization Policies | KNOMAD. Retrieved 5 July 2021, from https://www.knomad.org/publication/internal-migration-india-integrating-migration-development-and-urbanization-policies

20. Das, K. C., & Saha, S. (2013). Inter-state migration and regional disparities in India. online http://iussp. org/sites/default/files/event_call_for_papers/Inter-state% 20migration_ IUSSP13. pdf (accessed 15 March 2015).
21. Rani, S. (2016). Analysing Lee's Hypotheses of Migration in the Context of Malabar Migration: A Case Study of Taliparamba Block, Kannur District. Thematics Journal of Geography, 1(1).
22. Wang, Z., Song, K., & Hu, L. (2010). China's largest scale ecological migration in the three-river headwater region. Ambio, 39(5), 443-446.
23. Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. Population and development review, 431-466.
24. Covid 19 crowdsourced data, Retrieved 5 July 2021, from https://www.Covid19india.org/
25. Ashfan, Y., (2021). Threat of second wave of COVID-19 still exists: Experts, Retrieved 5 July 2021, from https://www.thehindu.com/news/national/karnataka/threat-of-second-wave-of-Covid-19-still-exists-experts/article33805526.ece
26. Nornha, G., (2021) India's Covid-19 resurgence unlikely to result in second wave, no impact on medium term outlook: Nomura(2021), Retrieved 5 July 2021, from India's Covid-19 resurgence unlikely to result in second wave, no impact on medium term outlook: Nomura - The Economic Times (indiatimes.com)
27. Check Train Coach Position, Retrieved 5 July 2021, from https://www.trainman.in/coach-position
28. Empowering Migrant Workers through Skill Development and Livelihood Generation(2020) Retrieved 5 July 2021, from http://niti.gov.in/sites/default/files/2020-11/MigrantsWorker_Inside_10.pdf