Interstate Development and Disparity

An Unsupervised Learning Approach

ADVAIT MOHARIR, RAJENDRAN NARAYANAN

The relationship between inequality and social welfare is an important yet understudied one. Using state-level data, a critical replication of a study conducted in the *Economic Survey* in 2021 is undertaken and it is found that the claimed positive inequality–welfare correlation does not hold. Then, unsupervised learning methods—principal component analysis and hierarchical clustering are used to classify states into groups based on a gamut of welfare indicators. It is found that some clusters form neatly around geographical divisions (north versus south) and confirm well-established developmental facts; significant heterogeneity among other regions (North-East) that do not adhere to established narratives is documented; and the development gap between states is found to be persistent and path-dependent. Finally, the need to strengthen rights-based legislations and establish a set of universal basic rights to course-correct for persistent inequalities is underscored.

The views expressed are personal and do not reflect the views of the institutions the authors are part of.

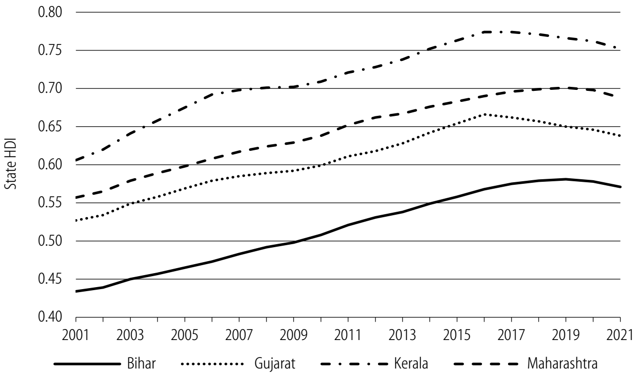
Advait Moharir (*advaithemant@gmail.com*) is an independent researcher. Rajendran Narayanan (*rajendran.narayanan@apu.edu.in*) teaches at Azim Premji University, Bengaluru and is affiliated with LibTech India.

There was much sound and fury concerning the economy and development in the recently concluded general elections in India. Inequality and poverty were hotly debated with no clear consensus. Two months prior to the elections, the union government released a "fact sheet" of the Household Consumption Expenditure Survey (HCES) 2022–23. Analysis based on the fact sheet as well as subsequent microdata suggests a decline in poverty as shown by Bhalla and Bhasin (2024). However, this enthusiasm is met with scepticism and intrigue by some other scholars. For instance, absolute consumption levels remain shockingly low with 34% of the population living on less than ₹100 a day (Anand 2024). Notwithstanding concerns about comparability with earlier rounds of the survey, Ghatak and Kumar (2024) point out how the poverty estimates from this survey do not square with other economic parameters such as wage stagnation, unemployment levels and inadequate industrial development. Using the fact sheet of 2022–23 HCES data and the Consumption Expenditure Survey data from 2011–12, there is a uniform Lorenz dominance of 2022–23 over 2011–12 indicating a uniform reduction in inequality in the last decade (Subramanian 2024). Given the disruptions caused by the pandemic and with growth reversals, the author is surprised to observe a first-order stochastic dominance of consumption across rural and urban areas between these two time periods suggesting a decline in poverty.

The unit-level data for the 2022–23 round of HCES was released after the conclusion of elections and this is likely to keep analysts occupied on issues of comparability with earlier rounds and excavate the extent of poverty and inequality in the country. While an all-India picture is immensely vital, it is important to delve into the interstate metrics owing to high levels of heterogeneity in incomes, growth and other development metrics.

There is belief among a section of economists that inequality is not an issue to be concerned about. Indeed, a chapter in the *Economic Survey* in 2021 claimed that reducing inequality is not important to improve socio-economic conditions across states. Arvind Panagariya, the chairperson of the Sixteenth Finance Commission, recently pitted inequality against poverty and argued that lowering systemic inequality is not a desirable goal for development (Panagariya 2024), and referred to those concerned about inequality as "inequality alarmists." However, this philosophy seems to be at odds with the findings of some other recent studies. Using the CPHS data between 2014 and 2019, Sahasranaman and Kumar (2023) demonstrate that states with higher income inequality have the lowest per capita

Figure 1: Trends in Human Development Index for Select Indian States



Source: Global Data Lab.

income for the bottom deciles, and vice versa. Importantly, they also highlight the higher stickiness of poverty among the poorer sections (first decile) in the more unequal states. There have also been huge debates surrounding the political economy between the northern and southern states. Andy Mukherjee (2024) compares income disparity and development metrics across the northern and southern states. He observes that the per capita domestic product in six northern states is much less compared to the five southern states. While nearly half of those in the age group of 18–23 are in higher education in Tamil Nadu, it is just one in six for the corresponding age group in Bihar. A report titled "Ten Years of NDA—A Guarantee Check" by the civil society platform Bahutva Karnataka, based on the analysis of the National Sample Survey data from 2012 and Periodic Labour Force Survey (PLFS) data up to 2022–23, reveals alarming levels of wage stagnation and stark interstate disparity. The entire household earnings of nearly half or more of the households in Chhattisgarh, Uttar Pradesh, Bihar, and Madhya Pradesh are less than the recommended national minimum wage of ₹375 per day (Bahutva Karnataka 2024).[1] In the south, the household incomes of at most one in five or one in six households are less than this threshold. These numbers are particularly disturbing, given that real per capita gross domestic product (GDP) has increased by 35% between 2012 and 2022. The southern states have historically invested extensively in building and enhancing essentials like health, education, nutrition and livelihood compared to the north. There is evidence to suggest that these have resulted not only in higher incomes but also a better quality of life in the south compared to the north. Analysing gross state domestic product (GSDP) from 1993–94 to 2019–20, Sinha et al (2023) find a positive correlation between economic growth and inequality. The authors have also categorised states according to low, medium and high inequality as measured by the Gini coefficient.

In this paper, we first debunk the statistically flawed claims made in the *Economic Survey* 2021. Using the same variables used in the *Economic Survey* 2021, we go beyond the north-south divide and embark on a data-based understanding of where all the states stand vis-à-vis their development profiles. Our paper is divided into three sections. In the first section, we briefly discuss why inequality matters, emphasising the reasons to lower it, and review how Indian states fare. In the next section, we present an alternate method of studying interstate disparities in development using unsupervised statistical learning methods. We use principal component analysis (PCA) followed by a hierarchical clustering algorithm to find states exhibiting similar patterns of development. Our paper can be situated as one possible extension of the paper by Sinha et al (2023). While they classify states based on inequality through temporal changes in the post-liberalisation period, we look at a more contemporary snapshot of interstate development profile by incorporating more development metrics. We end with some concluding remarks.

Why Inequality Matters

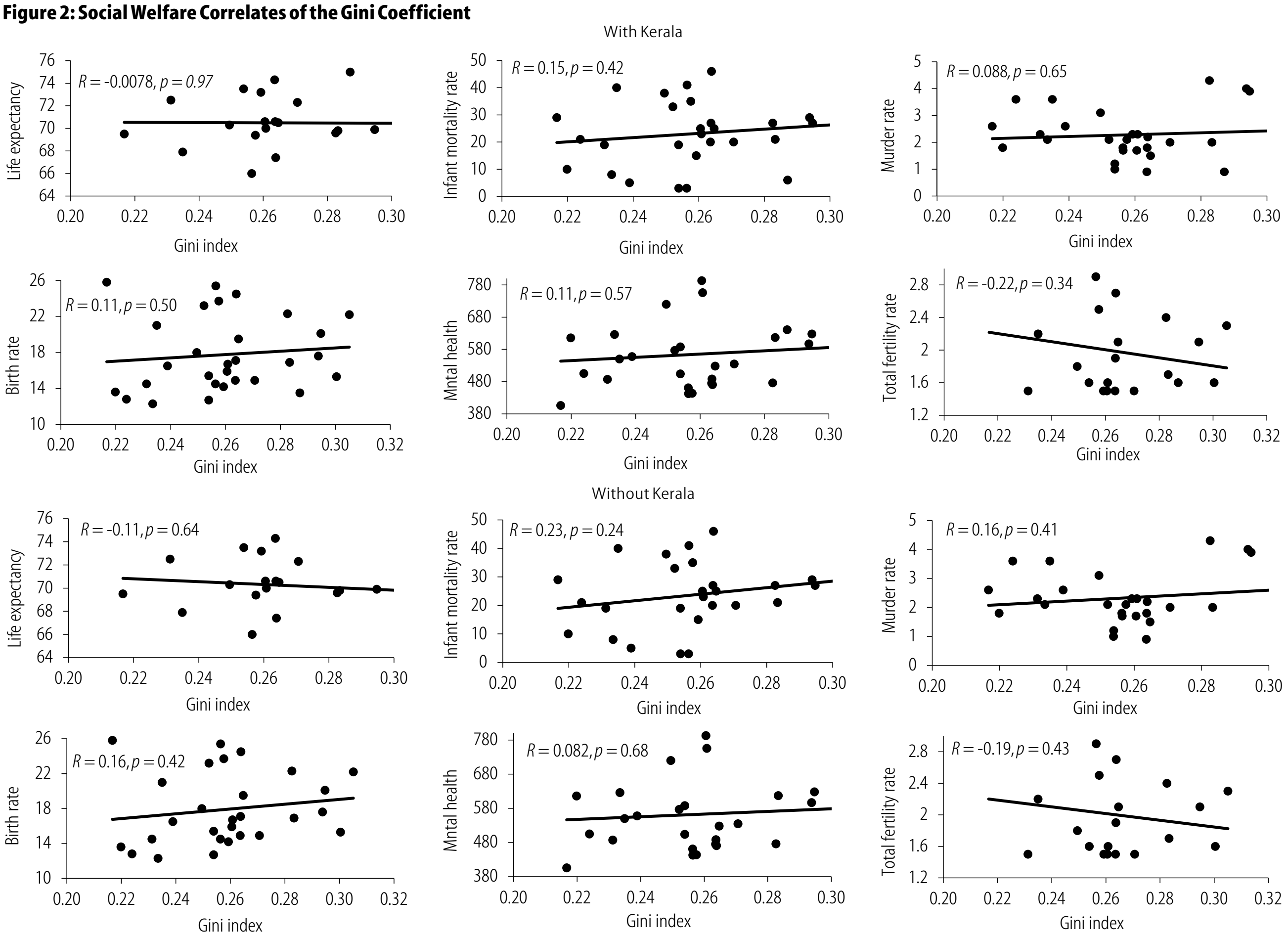
The extant literature has explored the effects of inequality of various political and economic outcomes. In particular, high inequality is associated with lower labour income (Autor et al 2020), and a diversion of resources from education (Galor and Moav 2004), as well as increased political polarisation and conflict (Stewart et al 2020; Esteban and Ray 2011). Over the last few years, inequality has been falling between countries, but rising within countries (Milanovic 2024), making it a crucial topic for policy consideration today. A recent working paper by researchers of the World Inequality Lab demonstrates that wealth concentration among the top 1% in India has been more pronounced between 2014–15 and 2022–23 and has reached historically high levels (Bharti et al 2024), exceeding those reached during the colonial era. There are also ethical, constitutional, and legal reasons to care about inequality, as outlined in Moharir and Narayanan (2024).[2]

The debate on development in India has centred around whether economic growth alone is sufficient (Bhagwati and Panagariya 2013), or whether the provisioning of social welfare is essential (Drèze and Sen 2013). It is argued that poverty reduction has been much slower compared to growth (Sen and Himanshu 2004). Recent evidence suggests that India's services-led growth has led to higher living standards, but these gains were skewed to high-income urban Indians (Fan et al 2023). Furthermore, locational inequality is a key determinant of living standards; specifically, almost a third of the variation in consumption can be explained by the state and the region (urban/rural) that an Indian lives in (Kumar et al 2022). Finally, studies also note a lack of convergence in growth between states, that is, on average, poorer states still grow slowly relative to their richer counterparts (Lamba and Subramanian 2020; Das et al 2015).

How do Indian states fare in terms of social indicators? Two patterns stand out: One, Human Development Index (HDI) trends between states have been nearly parallel over the last two decades (Figure 1), indicating that the gap between them is unchanged. Bihar's HDI score in 2021 is lower than Kerala's score in 1990. This corroborates early analysis of regional imbalances (Deaton and Drèze 2002). Two, over the last five years, the HDI trends have remained flat or shown a downturn, indicating stagnancy or deterioration in outcomes.

A relevant question then becomes: What is the impact of inequality on interstate social outcomes?

Figure 2: Social Welfare Correlates of the Gini Coefficient



Source: *Handbook of Statistics on Indian States*, NITI Aayog, NCRB, *Lancet*, NSS HCES–2022.

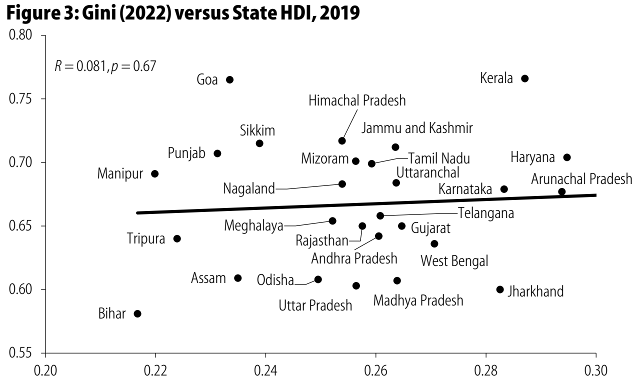
Economic Survey 2021

In its fourth chapter, the *Economic Survey* 2021 argued that there is a positive relationship between economic inequality and social outcomes. To substantiate this claim, for 29 states, the authors chose 10 key indicators of social welfare, such as life expectancy, drug use, education and health index, infant mortality, etc, and plotted them against the corresponding Gini coefficient: a standard measure of inequality. Observing a positive correlation, the survey concluded: "Thus, unlike in advanced economies, in India, economic growth and inequality converge in terms of their effects on socio-economic indicators."

However, this had multiple flaws. We demonstrate this by replicating the analysis done in the *Economic Survey*, by updating the data for the welfare indicators up to 2019, and report their correlates with the Gini index in 2022. First, the claimed relationship was driven largely by an outlier state: Kerala. The bottom panel of Figure 2 highlights this well. The correlation coefficient drops from 0% to -11% and 11% to 8% for life expectancy and mental health respectively. Murder rate and infant mortality show an increase, going from 9% to 16% and 15% to 23%, with and without Kerala respectively. A replication using the original data from the *Economic Survey* was recently done finding the same flaw (Chatterjee 2021).[3]

Second, the positive relationship between inequality and social outcomes is not robust. Most of the data for the social welfare indicators used by the survey were from 2016/17. A lot of the results using data from 2016/17 do not hold when replicated using 2019 data. For instance, the relationship between infant mortality rate and inequality is weakly positive instead of negative, while that of murder rate and inequality is flat. Therefore, the positive relationship between welfare and inequality as claimed in the *Economic Survey* was largely due to chance.

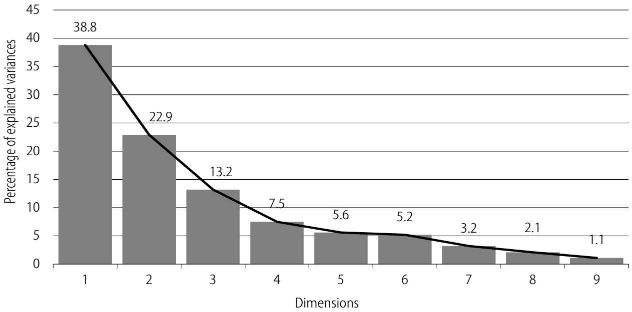
Figure 3: Gini (2022) versus State HDI, 2019



Source: NSS HCES 2022 and Global Data Lab.

The flawed analysis in the *Economic Survey* serves as a cautionary tale against relying on simplistic analysis when the number of observations are few, as this gives outsized power to outliers. This is demonstrated in Figure 3, which plots the Gini (2022) against HDI (2019). States like Kerala (which is highly unequal

Figure 4: Percentage of Explained Variation by Principal Component



Source: Authors' calculations.

but also developed), and Bihar (uniformly underdeveloped) can bias results significantly.

Clusters of Development

Very often we are confronted with statements such as "Kerala and Tamil Nadu are more developed" or that "the Gujarat model of development needs to be adopted for the whole country." Rhetorically powerful as they are, it would be useful to objectively assess where states stand vis-à-vis each other, based on some key development metrics. For this, we categorise states into clusters by exploiting the combined explanatory power of all the indicators used by the *Economic Survey*.

We use unsupervised statistical learning methods which are not used to predict but are used to explore and understand which data points are closer to each other and which ones are farther apart. We seek to find a way to combine the information contained in the variables measured for each state and establish which states exhibit similar patterns of development. Our method can be split into two steps. First, we do a PCA on the data to find low-dimensional summaries of high-dimensional data. We then perform a hierarchical clustering algorithm on the low-dimensional summaries to arrive at the clusters of states.

Principal Component Analysis

There are 9 variables in our data set; so our original data set consists of measurements in a 9-dimensional space and some of these variables are correlated with each other. PCA is a powerful technique that efficiently reduces the number of correlated dimensions into fewer uncorrelated dimensions, retaining the information content of the original set of variables. This is desirable because fewer dimensions are relatively more mathematically tractable. Mathematically, PCA works by doing an eigendecomposition of the covariance matrix of the variables. Each eigenvector—a linear combination of the variables—corresponds to a principal component direction and the corresponding eigenvalue shows the extent of variation captured in that principal component direction. The main outputs of PCA are: (i) the creation of new coordinate axes called principal components where most of the data lives and (ii) the values of each observation (states) measured in these new coordinate axes, known as PC scores.

PCA has some nice properties:

(i) Each principal component is a weighted linear combination of all the variables. The weights can be thought of as the contribution of the variables to that principal component direction.

(ii) The principal components (new coordinate axes) can be ordered in such a manner that the first few principal components account for a large proportion of variation in the original data. So the first principal component accounts for maximum variation, the second principal component accounts for the largest variability among what is remaining and so on.

(iii) A smaller number of principal components therefore acts as proxies for a large number of variables.

We construct the principal components using all the variables in the *Economic Survey* chapter, except fertility rate and life expectancy.[4] Additionally, we include the log of per capita net state domestic product (NSDP) to account for the role of income in development outcomes.

Figure 4, known as a scree plot, shows the percentage of variance accounted for by each principal component. We see that the first three principal components together capture 75% of the total variation. The trade-off of dealing with reduced dimensions (3 instead of 9) to build the clusters is that we are unable to capture the full extent of variation, but 75% of the total variation being captured is optimal enough.

Table 1 shows the first two PC scores (PC1 and PC2 henceforth) by state. To demonstrate how the states are separated on the PC dimensions, we select four states with the highest positive PC1 scores and four states with the most negative PC1 scores along with their per capita incomes. Table 2 shows which variables contribute most with PC1 and PC2. As a reminder, PC1 and PC2 axes are orthogonal to each other. By examining Tables 1 and 2 together, we can infer the relationship between the magnitude and sign of PC scores and the corresponding variables used to construct the PCs. For instance, the first four states in Table 1 have high *positive* PC1 scores and relatively low per capita incomes. From Table 2, we see that the variables infant mortality rate, birth rate, and murder rate contribute most positively to a high PC1 score. Additionally, the per capita income, health and education index contributes negatively to PC1. We can then infer that states with *high* positive PC1 scores have high infant mortality rates and birth rates, along with low income and bad overall health while those with high negative PC1 scores (bottom four states in Table 1) have low infant mortality, birth rates and good overall health. In this sense, on the PC1 dimension, Uttar Pradesh, Bihar, Madhya Pradesh, and Chhattisgarh tend to behave similarly while Mizoram, Kerala, Goa, and Sikkim exhibit a similar pattern.

Examining PC2 scores in Table 2 tells us that mental health, death rate and education contribute positively,

Table 1: PC1 and PC2 Scores for Selected States

State	Per Capita NSDP	PC1	PC2
Uttar Pradesh	43,061	3.57	–0.50
Bihar	26,978	3.27	–1.21
Madhya Pradesh	60,462	3.19	–0.25
Chhatisgarh	76,749	2.50	–0.70
Sikkim	2,48,691	–1.85	–2.20
Goa	3,13,973	–2.13	0.77
Kerala	1,47,951	–2.73	2.38
Mizoram	1,30,741	–2.87	–2.79

Source: Authors' calculations.

Table 2: PC1 and PC2 Scores by Variable

Variable	PC1	PC2
Infant mortality rate	0.91	0.28
Birth rate	0.90	–0.01
Murder rate	0.47	–0.09
Death rate	0.38	0.79
Drug use	–0.39	–0.80
Education index	–0.39	0.52
Mental Health Index	–0.46	0.55
Log (NSDP per capita)	–0.67	0.27
Health index	–0.73	0.25

Source: Authors' calculations.

while drug use contributes negatively. What these imply is that states like Kerala and Goa tend to have good mental health and education, along with relatively high death rates. High death rates do not mean that more people are dying. It just means that the lack of contribution of birth rate (-0.01) implies that these are states with low fertility rates. States like Sikkim, Mizoram, and Bihar having high negative PC2 scores suffer from high drug use.

Figure 5 provides a visual summary of the previous results and shows the values of each state (PC scores) in the new coordinate axes, PC1 and PC2. The colour scale is a proxy for the correlation between the two (high to low). The top left quadrant (second quadrant) consists of states with negative PC1 scores and positive PC2 scores. These are states scoring high on income and development outcomes and comprise mainly states from the south, west and West Bengal. The bottom left quadrant (3rd quadrant) has states with negative PC1 and PC2 scores. These states, while having a high per capita income and performing well on many development outcomes, suffer from high drug use, and consist mainly of north-eastern states and Punjab. The final bunch of states to the right of the new Y-axis (PC2) have all positive PC1 scores, indicating poor performance on key development indicators. However, there is a mix of positive and negative PC2 scores, indicating some heterogeneity in drug use and death rates among others. These states mainly consist of north and east Indian states.

The visual map in Figure 5 lends easily to think of states as being clustered together. This can be made more precise by performing a hierarchical clustering analysis on the PC scores, which we do next.

Hierarchical Clustering

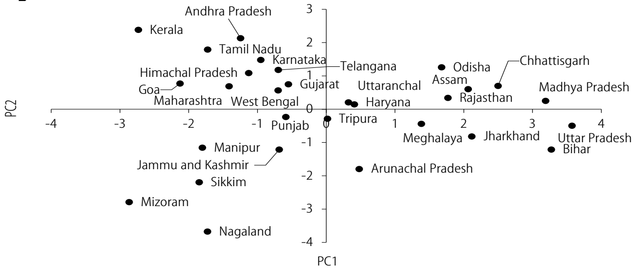
Cluster analysis is used to find non-overlapping sub-groups in a data set such that observations within a sub-group are similar and observations across sub-groups are dissimilar. Each sub-group is called a cluster. There are several algorithms for clustering such as model-based clustering and K-Means. We use a method called the hierarchical clustering algorithm which is non-model based and proceeds in a bottom-up manner.

We cluster the n=29 observations (states) based on their first 3 PC scores. In hierarchical clustering, we start with n unique clusters, called leaves, and proceed sequentially to merge similar leaves to form branches and then fuse similar branches to form a tree. In our data set, the algorithm starts with each state as a separate cluster and the algorithm terminates by creating 1 cluster composed of all the states. Both these are not informative as the former indicates that all states are different while the latter suggests that all the states are similar. There are optimal or suboptimal sweet spots with fewer branches (clusters) such that all the leaves within a branch are similar while the leaves in different branches are dissimilar. For this, we need to cut the tree and create branches (clusters) using the following steps (James et al 2021):

(1) Start with n observations and a choice of distance measure such as Euclidean distance or Manhattan distance and treat each observation as its own cluster.

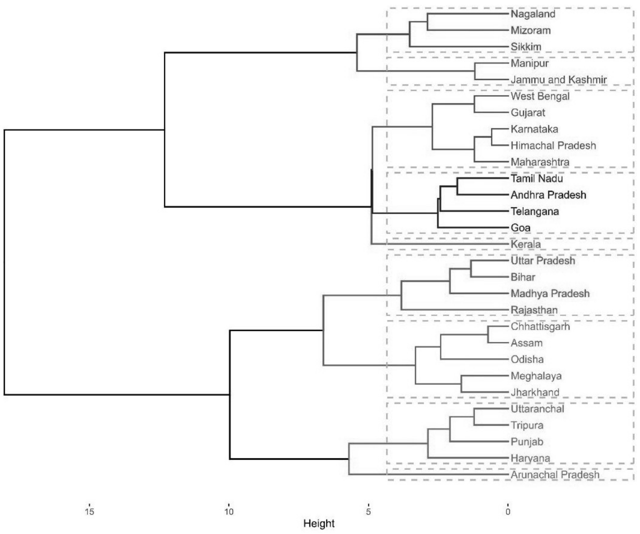
(2) For $j = n, n\text{-}1, \ldots, 2$:

Figure 5: PC1 versus PC2 for All States



Source: Authors' calculations.

Figure 6: Hierarchical Clustering of Indian States (k=9)



Source: Authors' calculations.

(a) Compute all pairwise distances between the j clusters using the chosen distance measure. Fuse those clusters which are closest to each other.

(b) Compute the new pairwise inter-cluster distance measure among the j -1 remaining clusters.

For distance, we use the Manhattan distance, which is a robust choice because it avoids giving undue importance to observations that are far away from the rest of the data. Manhattan distance is given by the sum of the absolute differences between two points. Having chosen a distance metric, one needs a way to fuse branches or existing smaller clusters to form new clusters. The single linkage function picks the nearest neighbours of two clusters and computes all the pairwise distances. The complete linkage function picks points in two clusters that are farthest from each other while the average linkage function, as the name suggests, computes the average of the data points within any two clusters and fuses those that are closest. For the linkage function, we choose Ward's method which chooses clustering steps such that the increase in the error sum of squared (ESS) is minimised and tends to create relatively more equi-sized clusters.

Figure 6, known as a dendrogram, shows the results for K=9 branches (clusters). The X-axis in Figure 6 shows the height at which the states are fusing to form clusters and the y-axis shows the state names. How do we interpret this?

The algorithm calculates pairwise distances between all the states based on the 3 PC scores and finds those states which are closest to each other. Observe in Figure 6 that the height at which Chhattisgarh and Assam merge and the height at which Karnataka and Himachal Pradesh merge are the lowest. Among all the possible pairs that can be formed, Chhattisgarh pairs best with Assam while Karnataka pairs best with Himachal Pradesh. Subsequently, Gujarat pairs best with West Bengal. At each stage, a state can join an existing branch or two branches can fuse to form a bigger branch. For example, Maharashtra joins the branch of Karnataka and Himachal Pradesh suggesting that Maharashtra is most similar in its development profile to the pair of states, Karnataka and Himachal Pradesh, rather than with any other individual state. The algorithm proceeds sequentially. The vertical dotted line in Figure 6 depicts where the tree is cut to show a meaningful set of clusters. An important caveat is that two clusters cannot be ordinally compared, that is, we cannot say that one cluster is better or worse than another cluster.

We discuss the results generated using 9 clusters. Each cluster will have cliques within it suggesting that even within a cluster there are more similar sub-groups. The first cluster (from the top) comprises mainly the north-eastern states. These are the same states scoring high in the drug index, as well as some development outcomes. Nagaland and Mizoram appear to be paired best in this regard. Manipur and Jammu and Kashmir form the next cluster.[5]

The third cluster from the top is perhaps the most interesting. Within this cluster, one clique is the trio of Karnataka, Himachal Pradesh, and Maharashtra while the other clique comprises West Bengal and Gujarat. One could characterise this as either "West Bengal is as good as having Gujarat's development model" or "Gujarat is as bad as West Bengal." Proceeding in this manner we can infer that Andhra Pradesh and Tamil Nadu are a clique and that Telangana is the next to join this clique followed by Goa. Kerala is far away from all the cliques in terms of development metrics and income and is assigned its own cluster, as is Arunachal Pradesh.

The bottom four clusters consist of states performing poorly on development indicators. It is on expected lines to see that Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh form a cluster giving a stamp to the cliche BIMARU states. Another cluster comprises the eastern states like Odisha, Chhattisgarh, and Jharkhand, which also happen to have a higher proportion of tribal population compared to other states in mainland India. And these states are most similar, not to their neighbours, but to north-eastern states like Assam and Meghalaya. This is true also for the second last cluster consisting of smaller north Indian states, where Uttarakhand lies on the same branch height as Tripura. This indicates that geography, while being a reasonable heuristic for predicting broad developmental trajectories, is limited in explaining more granular similarities. An important feature of note is that the last four clusters remain unchanged, when we choose to classify with six (Appendix Figure A1, p 80) or nine clusters, indicating that they are robust.

Concluding Remarks

In this paper, we implement two empirical exercises. One, we take stock of where India's states are vis-à-vis some standard welfare measures, and debunk the spurious claim made by the *Economic Survey* that inequality and welfare are positively correlated. Two, we provide an alternative way of examining differences between states using unsupervised statistical learning methods. In doing so, we identify various clusters of development. What can we infer from these exercises?

One, we document significant heterogeneity among states. While some of these clusters form neatly around geographical lines (north versus south) and conform to long-established stylised facts, there is an interesting set of intermediate outcomes that resist easy classification. For instance, the north-eastern states are spread across four clusters, and some of them are more similar to mainland states than to each other. States such as Gujarat and West Bengal, which have followed very different models of development, are closest together. This indicates that one needs to move beyond geographical and institutional explanations to get a better grasp of varied disparities among states.

Two, the development gap among states is persistent and the trajectory is path-dependent. As shown in Figure 1, states starting on high/low development paths remain there and do not converge. The southern states continue to reap the benefits of a relatively less extractive colonial regime (Banerjee and Iyer 2005), as well as massive investment in infrastructure, health and education done post-independence, relative to their northern counterparts. This also means that interstate inequality is not self-correcting and needs active intervention. Additionally, it is important to note that high growth, while necessary to alleviate within-state poverty, is not sufficient in reducing the between-state gap. For example, excluding the COVID-19 year (2020–21), both Kerala and Tamil Nadu's real gross state domestic product (GSDP) has grown at the same average rate of 6.8% over the last 10 years. Yet they remain on the opposite ends of the HDI spectrum.

Three, the lack of frequent, granular and administrative data is a major binding constraint for research and policy. The gap between the last two representative household consumption surveys was 12 years. The census, due in 2021 and pushed due to COVID-19 and other exigencies, continues to be delayed. Furthermore, many of the metrics used to construct other key statistics like GDP are due for an update. An important aspect of good policymaking is careful and rigorous statistical analysis, which in turn is only as good as the underlying data. Given that metrics like employment are now being measured every quarter (PLFS), the state must start measuring and reporting consumption, income, poverty and wealth numbers regularly.

Finally, economic disparities significantly undermine social solidarity and civic cooperation. The need is to bolster existing rights-based legislations on education, employment, food security, etc, and ensure that health and pensions become rights too, among other things. In 2023, the Rajasthan government enacted a right to health and a minimimum income guarantee law under which an urban employment programme and inflation-adjusted pensions became a right. These are welcome steps and can play a vital role in absorbing some amount of distress shocks. Expanding this base and establishing unimpeachable universal basic rights are crucial steps in reducing interstate inequalities in the future.

NOTES

1 This recommended minimum daily wage was made by the Anoop Satpathy Committee in 2019 which was set up by the Ministry of Labour and Employment. However, the recommendations have been ignored.

2 The authors argue for progressive taxation such as wealth tax and inheritance tax on the super elite as a source of funds to smoothen persistent inequality.

3 For another replication using the original data see https://twitter.com/MoharirAdvait/status/1357297752372350977.

4 We drop these variables as data for this is reported only for 22 of the 29 states.

5 While the development profile of most states would remain similar with newer data, we posit that given the conflict in Manipur, we anticipate many changes in the development profile of Manipur.

REFERENCES

Anand, Ishan (2024): "What Does the Data from the Household Consumer Expenditure Survey 2022–23 Tell Us?" *The India Forum*, 10 July, https://www.theindiaforum.in/economy/poverty-india-over-last-decade.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson and John Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics,* Vol 135, No 2, pp 645–709, https://doi.org/10.1093/qje/qjaa004.

Bahutva Karnataka (2024): "Ten Years of NDA · A Guarantee Check—Bahutva Karnataka," Bahutva Karnataka, https://bahutvakarnataka.wordpress.com/guaranteechecks/.

Banerjee, Abhijit and Lakshmi Iyer (2005): "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review,* Vol 95, No 4, pp 1190–1213, https://doi.org/10.1257/-0002828054825574.

Bhagwati, Jagdish and Arvind Panagariya (2013): *Why Growth Matters: How Economic Growth in India Reduced Poverty and the Lessons for Other Developing Countries*, PublicAffairs.

Bhalla, Surjit and Karan Bhasin (2024): "Poverty in India Over the Last Decade," *Economic & Political Weekly*, Vol 59, No 28, https://www.epw.in/journal/2024/28/special-articles/poverty-india-over-last-decade.html.

Bharti, Nitin, Lucas Chancel, Thomas Piketty and Anmol Somanchi (2024): "Economic Inequality in India: The Rise of the Billionaire Raj," World Inequality Lab Working Paper 2024/09, World Inequality Lab.

Chatterjee, Shiladitya (2021): "The Role of Inequality and Growth in Determining Social Outcomes," *Economic & Political Weekly*, Vol 56, No 30, https://www.epw.in/journal/2021/30/commentary/role-inequality-and-growth-determining-social.html.

Corak, Miles (2013): "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives,* Vol 27, No 3, pp 79–102, https://doi.org/10.1257/jep.27.3.79.

Das, Samarjit, Chetan Ghate and Peter Robertson (2015): "Remoteness, Urbanization, and India's Unbalanced Growth," *World Development,* Vol 66, No C, pp 572–87.

Deaton, Angus and Jean Drèze (2002): "Poverty and Inequality in India: A Re-Examination," *Economic & Political Weekly,* Vol 37, No 36, pp 3729–48.

Drèze, Jean and Amartya Sen (2013): *An Uncertain Glory: India and Its Contradictions*, New Jersey: Princeton University Press.

Esteban, Joan and Debraj Ray (2011): "Linking Conflict to Inequality and Polarization," *The American Economic Review*, Vol 101, No 4, pp 1345–74.

Fan, Tianyu, Michael Peters and Fabrizio Zilibotti (2023): "Growing Like India—The Unequal Effects of Service-Led Growth," *Econometrica,* Vol 91, No 4, pp 1457–94, https://doi.org/10.3982/ECTA20964.

Galor, Oded and Omer Moav (2004): "From Physical to Human Capital Accumulation: Inequality and the Process of Development," *The Review of Economic Studies,* Vol 71, No 4, pp 1001–26.

Ghatak, Maitreesh and Rishabh Kumar (2024): "Poverty in India Over the Last Decade: Data, Debates, and Doubts," *The India Forum*, 10 April, https://www.theindiaforum.in/economy/poverty-india-over-last-decade.

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani (2021): *An Introduction to Statistical Learning: With Applications in R*, New York: Springer US, https://doi.org/10.1007/978-1-0716-1418-1.

Kumar, Rishabh, Sriram Balasubramanian and Prakash Loungani (2022): "Inequality and Locational Determinants of the Distribution of Living Standards in India," *Structural Change and Economic Dynamics,* Vol 61, June, pp 59–69, https://doi.org/10.1016/j.strueco.2022.02.007.

Lamba, Rohit and Arvind Subramanian (2020): "Dynamism with Incommensurate Development: The Distinctive Indian Model," *Journal of Economic Perspectives,* Vol 34, No 1, pp 3–30, https://doi.org/10.1257/jep.34.1.3.

Milanovic, Branko (2024): "The Three Eras of Global Inequality, 1820–2020 with the Focus on the Past Thirty Years," *World Development,* Vol 177, May, https://doi.org/10.1016/j.worlddev.2023.106516.

Moharir, Advait and Rajendran Narayanan (2024): "An Inheritance Tax Will Help Reduce Inequality," *Hindu*, 7 May, https://www.thehindu.com/opinion/op-ed/an-inheritance-tax-will-help-reduce-inequality/article68149271.ece.

Mukherjee, Andy (2024): "Why India's South Rejects Modi—and Why It Matters," *Deccan Herald*, 8 April, https://www.deccanherald.com/opinion/why-indias-south-rejects-modi-and-why-it-matters-2969076.

Panagariya, Arvind (2024): "Don't Lose Sleep over Inequality," *Times of India*, 2 April, https://timesofindia.indiatimes.com/blogs/toi-edit-page/dont-lose-sleep-over-inequality/.

Sahasranaman, Anand and Nishanth Kumar (2023): "Distributional Dynamics of Income in Indian States," *Economic & Political Weekly,* Vol 58, Nos 25–26, https://www.epw.in/journal/2023/25-26/special-articles/distributional-dynamics-income-indian-states.html.

Sen, Abhijit and Himanshu (2004): "Poverty and Inequality in India: I," *Economic & Political Weekly,* Vol 39, No 38, pp 4247–63.

Sinha, Manjisha, Sendhil Ramadas and P Ramasundaram (2023): "Are Spectacular Growth and High Inequality Two Sides of the Same Coin?," *Economic & Political Weekly,* Vol 58, No 19, https://www.epw.in/journal/2023/19/special-articles/are-spectacular-growth-and-high-inequality-two.html.

Stewart, Alexander J, Nolan McCarty and Joanna J Bryson (2020): "Polarization under Rising Inequality and Economic Decline," *Science Advances*, Vol 6, No 50, eabd4201, https://doi.org/10.1126/sciadv.abd4201.

Subramanian, S (2024): "The Household Consumption Expenditure Survey 2022–23," *The India Forum*, 3 April, https://www.theindiaforum.in/economy/household-consumption-expenditure-survey-2022-23.

Appendix

Figure A1: Hierarchical Clustering of Indian States (k=6)