Towards Building a District Development Model for India Using Census Data

Dibyajyoti Goswami
Indian Institute of Technology
Delhi, India
10dibyajyoti@gmail.com

Shyam Bihari Tripathi
Indian Institute of Technology
Delhi, India
shyam.tripathi7@gmail.com

Sansiddh Jain
Indian Institute of Technology
Delhi, India
sansiddh@gmail.com

Shivam Pathak
Indian Institute of Technology
Delhi, India
shivam.pathak.cer11@iitbhu.ac.in

Aaditeshwar Seth
Indian Institute of Technology
Delhi, India
aseth@cse.iitd.ernet.in

ABSTRACT

Models for socio-economic development are useful for planners to build appropriate policies. Such models are ideally constructed based on empirical data, and we take up the problem of working towards a district development model for India by using two waves of census data. India has almost six hundred districts, diverse in terms of their social and economic development, and hence presents a unique natural experiment to understand how social and economic factors interplay with one another. In this paper, we present some interesting observations we are able to make from the analysis of census data from the years 2001 and 2011, and also raise some questions calling for the need for additional ethnographic and other surveys to be able to understand the underlying mechanisms that would have led to the observed patterns.

KEYWORDS

social development, economic growth, employment, health, literacy, industrialization, census

ACM Reference Format:

Dibyajyoti Goswami, Shyam Bihari Tripathi, Sansiddh Jain, Shivam Pathak, and Aaditeshwar Seth. 2019. Towards Building a District Development Model for India Using Census Data. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '19), July 3–5, 2019, Accra, Ghana.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3314344.3332491

COMPASS '19, July 3–5, 2019, Accra, Ghana

ACM ISBN 978-1-4503-6714-1/19/07…$15.00
https://doi.org/10.1145/3314344.3332491

1 INTRODUCTION

There has been a wide ranging debate on the interplay between economic growth and social development, most recently discussed in exchanges between Jagdish Bhagwati and Amartya Sen [32]. The arguments on one side favour a model of economic growth as a precursor to social development, with little concern about growth in inequality which is believed to be a temporary and necessary phenomenon before the gains of economic growth get more widely distributed. The arguments on the other side place a greater emphasis on social development indicators such as health and education in the context of a liberal democracy, as being necessary to create opportunities for people to realize their individual capabilities, eventually leading to economic growth as well. These somewhat conflicting perspectives lead to different views on policy and resource allocation. The economic growth camp favours policies that encourage industrialization, corporate subsidies, tax breaks, etc to accelerate economic growth which if further left unfettered from government regulations through free market mechanisms will lead to natural processes of more widespread economic growth and social development. The social development camp favours welfare mechanisms and re-distributive policies to help the poor improve their living conditions and grab growth opportunities, while also having access to safety nets so as to not fall back into poverty when faced with income or health shocks.

This is a complex debate not only because evidence for one or the other approach being better might be lacking, but also because both approaches respect different underlying values [13]. We cannot possibly resolve the debate but we hope that being able to understand how the six hundred different districts of India have grown over the years, will help reveal some patterns that could inform the debate. We start with answering questions such as the following. Is economic growth happening equitably across the country? Are economic growth and social development going hand in hand? Does the government seem to have favoured one approach over the other? To answer such questions, we use data from the Indian census for 2001 and 2011, and study the changes that have happened in different economic and social indicators at the district level.

We find a strong relationship between literacy rates in districts, and employment in non-agricultural sectors (manufacturing and services) in the districts. This is expected since these industry sectors would prefer educated people. We also find that development indicators related to important amenities such as the use of cleaner sources of fuel, or better construction material used in houses, also improve with an increase in non-agricultural employment of the households. This shows that non-agricultural employment generates more disposable income than agricultural employment, which households are then able to plough into improving living standards for themselves. However, we also find that the manufacturing and services sectors have not expanded to other geographies, rather they have grown in scale in the same districts, leading towards increasing inequality among districts.

Another interesting set of observations we make is that households seem to prefer acquiring assets more than investing in essential amenities such as the use of cleaner sources of fuel for cooking. The government also seems to have preferences in the investment it makes in social infrastructure, and we find that electrification and lighting has seen faster change than provisioning of sources of drinking water. However these investments in social infrastructure do not seem to influence people's behaviour on acquiring assets or essential amenities for healthier living conditions.

We also find a significant drop in female employment, caused by a withdrawal of women from casual employment but without a commensurate increase in regular employment. Such a reduction in the movement of women from home to the workplace has been associated historically with an increase in gender inequality.

Our key contributions in this paper are to point out such observations regarding the interplay between social and economic development indicators. While some of our observations are new and some have been made by other researchers as well, our method of formulating hypotheses using easily explainable categories is novel and is aimed at making observations on macro socio-economic changes more accessible to people. Further, our analysis is conducted at the district level using census data, whereas most other work on similar indicators has been done only at the state or national level using surveys that have much smaller sample sizes than censuses.

When presenting these patterns seen in the economic growth and social development happening across the country, we also highlight many questions that are left unanswered but which lead to the differing value-based concerns between the two approaches to economic growth and social development. For example, many complex dynamics are generated as a result of unequal spatial distribution of manufacturing and services sectors, mentioned above. It is seen that this concentration of employment growth in existing hubs of employment, without a spatial expansion into other districts, results in people migrating to urban centers for employment when they are not able to find employment locally in rural areas [15]. The underlying cause for a stress in rural employment arises because agriculture as the primary source of rural employment is no longer able to absorb more people due to reduced sizes of inherited landholding as families grow, and agriculture also faces diminished income due to pro-consumer policies that end up being farmer-unfriendly [4, 19]. Much of this migration is also seasonal, which serves as the conduit for income generated in urban areas to flow back to rural areas [14, 22], leading to the development of these areas indirectly via this urban income rather than directly through local employment generation. Further, this pattern of indirect development via rural-urban migration is also known to be exploitative because migrant workers do not hold as strong a bargaining power to assert their rights and hence get lower wages [34].

Similarly, there could be many reasons behind households preferring to invest in assets and governments prioritizing electrification over drinking water. It is possible that assets such as television sets may be economically useful for the households in staying informed, or it may also be that advertising fueled consumerism encourages people to spend on non-essentials rather than make healthier lifestyle choices. Similarly, the government may see a stronger need to invest in electrification for immediate economic growth, rather than clean drinking water for healthier citizens which may translate into economic gains only in the long-term.

Such questions cannot be answered by the census data alone but are important concerns in choosing between the two approaches because they highlight which values are preserved or violated in the processes unfurled by different approaches. Whether to tolerate exploitation of the weak for possible eventual empowerment, or to guide people on how to spend their disposable income, or to provide facilities for work but not basic human needs, are the kind of concerns at the heart of the debate that arise when mechanisms that are producing the observed statistics are revealed.

Such limitations to understand the underlying dynamics behind broad correlations and anomalies exist not just with census data, but with the analysis of any big data sources [41]. In the field of development itself, sources like satellite imagery [20], commodity prices [7], cellphone records [6], etc are now being used actively to spot relationships between different variables, but need to be supplemented with other studies to understand the mechanisms that would be giving rise to the observations made from big-data analysis. As part of future work, we also outline our vision of building an economy monitor that combines observations from big-data sources with quantitative and qualitative data generated through participatory media platforms, which may help explain these observations. We feel that such a platform will help reveal richer insights to inform the debate between economic and social development, and eventually contribute towards building development models that have explicit values embedded within them. *We feel that unless these values are not explicitly debated, mere statistics alone will get interpreted in different ways to suit different approaches.*

2 RELATED WORK

Policy makers in India have traditionally relied on planned resource allocation towards economic and social development goals [10]. Our work is an empirical effort that can inform such planning activities by drawing out observations on how different types of districts in India have evolved so far, what gaps remain to be addressed, patterns that warrant further investigation, and outcomes likely to arise if the same model of development is followed. Other work with a similar empirical approach has relied on the use of satellite data to observe growth in inequality across different regions [18]. Extensive empirical work has also been conducted using five yearly sample surveys in India to understand patterns of change in female employment [36] and labour force participation [39]. Census data has been used to study the impact of district level administrative boundaries on access to household amenities [3]. A spatial concentration in economic activities is also evident using data from the Annual Survey of Industries [24, 26].

All such work including our own can benefit from further analysis to understand the underlying dynamics behind these observations, and lead to value-based criteria in building policies. Our eventual goal is to build such a district level model of development. Country level models for emerging economies to transition from an agrarian economy to manufacturing or services, has seen considerable discussion including an IMF working paper on the model followed by India [23]. Models for women participation in the workforce of first declining and then rising with development, have also been proposed [17, 28]. Models have not been proposed at the district level however, especially in the context of India, and which incorporate complex dynamics that can be judged within different value systems. Our vision of utilizing data from participatory media networks to uncover the mechanisms shaping and arising from the observed patterns, can help build rich models for district development.

3 METHODOLOGY & DATA PREPARATION

3.1 Overall methodology

The diversity of India presents a natural experimental setup to study how different districts have evolved over time. We use census data for two years separated by a decade, and select a mix of six social development and economic growth indicators, to examine the patterns of change over this decade. Around these six indicators, we construct six hypotheses and test these hypotheses. All these hypotheses are largely about comparing the pace of change of an indicator with respect to other intervening variables, to get a sense of which indicators move faster under which conditions. These are precisely the kind of empirical insights that can inform the Bhagwati-Sen debate about when social development happens viz-a-viz economic growth, and vice-versa. Our approach at a high-level is similar to association rule mining [1], but we choose to use a simpler method of comparing the mean rate of change in different indicators (at a statistically significant level) to come up with easily interpretable patterns. We also develop a unique discretization method for real-valued data across many categories, to make the hypotheses more interpretable.

In the following sections, we outline the data preparation steps we undertook to evaluate six hypotheses of social and economic change. We start with a description of the census data, followed by the discretization method, and then present examples of interesting patterns which we evaluate in detail in subsequent sections.

3.2 Census of India: 2001 and 2011

The Government of India conducts a population census every ten years. We used data from the 2001 and 2011 censuses, available from the official website [8]. The census reports the number of households in each spatial unit (village, district, state), belonging to 90 different categories including the type of construction of the house, the cooking fuel which is used, assets owned by the household, type of employment, sector of employment, etc. The 2001 data was available only at the district level, hence we did our analysis at the district level. Between 2001 and 2011, 47 districts split into smaller districts due to administrative changes, and for our analysis we grouped them back into the original 593 districts that existed as of 2001.

3.3 Discretization of variables

Each category variable provided by the census may have multiple parameters, and reports the number of households in a district for each parameter. For example, for the variable regarding the primary type of fuel used for cooking by households, the census reports separately the number of households in a district using firewood, kerosene, LPG (Liquid Petroleum Gas), PNG (Piped Natural Gas), biogas, etc. For our analysis, we wanted to reduce these to a single value for each variable and we developed the following procedure. We first group the mutually exclusive parameters within a variable into three broad parameter types of rudimentary (Rud), intermediate (Int), and advanced (Adv). For example, firewood is considered as a rudimentary type of fuel for cooking, kerosene and cow dung are grouped together as an intermediate type, and PNG, LPG and bio gas are grouped together as advanced types of fuel for cooking. This grouping is showing in Table 1.

Variable	Using/Access to	Level - 1	Level - 2	Level - 3
Bathroom Facility	Rud: No Latrine facility	65-82	20-40	18-40
	Int:Pit Latrine	0-5	30-45	0-10
	Adv :Piped Sewer/Septic Tank	15-28	25-40	50-70
Fuel for Cooking	Rud:Firewood	60-80	30-50	20-40
	Int:Cow Dung/Kersosene	5-15	40-60	5-20
	Adv:LPG/PNG/Bio gas	15-35	5-20	45-65
Condition of Household	Rud:Dilapidated House	5-10	0-5	0-5
	Int:Livable House	55-65	40-50	25-35
	Adv:Good House	30-40	45-55	65-75
Main Source of Light	Rud:No source of light	0-5	0-5	0-5
	Int:Kerosene oil/Other oils	70-80	30-50	5-15
	Adv:Electricity/Solar Light	20-30	50-70	85-95
Main Source of Water	Rud:Well/Spring/River	40-70	2-20	5-15
	Int:Hand Pump/Tube Well	2-25	55-80	10-28
	Adv:Tap Water/Treated water	20-40	10-28	60-85
Asset Ownership	TV	15-30	30-50	60-85
	Telephone	35-55	40-60	50-60
	2-Wheeler	5-12	5-18	20-40
	4-Wheeler	0-2	0-5	2-12

Table 1: Census variables to classify districts in terms of levels for different indicators: Shown is the % of households within various indicators

We then do a k-means clustering on the districts based on the percentage of households in each district that use different types of fuel: rudimentary, intermediate, and advanced. Figure 1 shows a box-plot for the distribution of districts across three levels (k = 3) in terms of their use of different types of fuel for cooking.

This allows us to label each district as a level-1/2/3 district: Level-1 districts predominantly use rudimentary types of fuel for cooking, level-2 districts primarily use intermediate types of fuel, and level-3 districts predominantly use advanced types of fuel for cooking. We follow the same method for other indicators also. Table 1 indicates the percentage of households having access to rudimentary, intermediate, and advanced measures of the indicators for each level. This

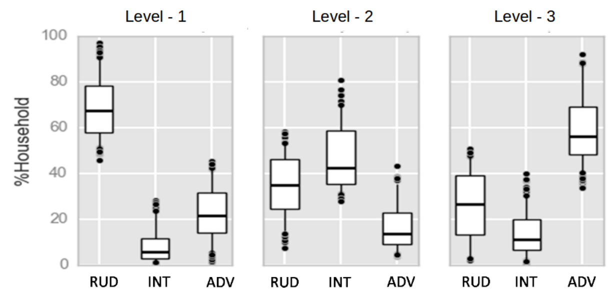

Figure 1: An example of discretizing the level of development of districts in terms of the fuel used for cooking in the district households. Shown is the distribution of the % of households using different types of fuel (rudimentary, intermediate, or advanced) across three district clusters. These clusters are used to label the level of a district for the fuel for cooking indicator

method therefore allows us to map each district to a single coarse value for each variable.

We experimented with different values of k for different variables in terms of the quality of clusters obtained, and eventually settled on k = 3 as a reasonable and uniform mapping of districts for all the variables. We justify this choice of k in the supplementary material [16], and also test the different hypotheses suggested in the paper for robustness with different values of k.

This method of discretization is useful for several reasons. First, as shown in the book *Factfulness* by Hans Rosling [35], who used a similar 4-level mapping for different stages of development of countries and regions, such a coarse mapping is easy for people to interpret and to easily compare different districts with one another. Second, it reduces the variables to a single quantity without assigning arbitrary weights to club together multiple parameters for each variable. Third, as we show next, it allows us to compare different variables with one another using simple probabilistic analysis to determine broad patterns, instead of more complex regression methods which may be hard to interpret and to determine significant relationships. This is the key method we use in this paper to study the relationships between different variables. Note that each district may be marked as belonging to a different level for different variables, for example, a district could be at level-1 in terms of its dominant use of fuel for cooking, but at level-3 in terms of asset ownership, and so on. Seeing how districts move from one level to the other across different variables, allows us to determine broad patterns about which variables tend to move first before others, and any inter-dependencies that might exist between the variables.

Variable	% of population employed in	District Type		
		Non Agri (in %)	Agri (in %)	High Unemployment (in %)
Employment	Unemployed	50-60	40-50	55-65
	Agricultural labour	5-10	25-35	15-20
	Non Agricultural work	22-35	10-15	8-15

Table 2: Census variable to classify districts in terms of type of employment: Shown is the % of population in different types of employment

We apply the same method to also classify districts in terms of the dominant type of employment: agricultural, non-agricultural, or high unemployment as given in Table 2. Continuous variables such as female employment and literacy are retained as such since they did not have multiple internal parameters.

3.4 Changes between 2001 to 2011

We need to allow for comparison of a district over the two census years, to determine whether the district moved from one level to another, for different variables. To do this, we first obtain the clusters using 2011 data since it showed more diversity in all the variables. For each variable, we then calculate the centroid for every cluster and determine the level of a district in 2001 by seeing which centroid is the closest for the district. This allows us to obtain Table 3 which shows the percentage of districts that moved from a lower level to a higher level between 2001 to 2011, stayed the same, or even dropped from a higher level to a lower level. This change is shown separately for districts based on their level as of 2001.

There are several interesting patterns to notice. Most level-3 districts for any variable show no-change (last column in Table 3), which is understandable because these districts were already at the highest level of development for that variable. There is a negative growth in some districts though, which we believe is due to significant inbound migration into these districts that caused an increase in the population living in slum areas and suburbs due to people coming in search of better employment opportunities to districts placed at higher levels of development. We also see that there has been an overall drop in female employment. Interestingly, there has been a significant drop for females in marginal employment (less than six months of work in a year), indicating either an increase in formalization in female employment or a decrease in overall female participation in the workforce. We will get back to this topic later in the paper.

Getting back to Table 3, we see that many districts have improved for some variables at level-1 and level-2, but fewer districts have improved for some other variables. Variables that require government investment in infrastructure, such as electrification to alter the main source of light used by households, or water pipelines and handpumps to alter the main source of water, show different degrees of change. This possibly indicates that different priorities may have been placed by the government on these two variables. In the same way, discretionary variables determined more by household decision making, also show different degrees of change. Districts at level-2 in asset ownership have improved significantly, but not districts at level-2 on the condition of household, for instance. In the following sections, we examine such patterns in more detail.

Variable	Level 1 (values in %)			Level 2 (values in %)			Level 3 (values in %)		
	+ve Growth	-ve Growth	No Change	+ve Growth	-ve Growth	No Change	+ve Growth	-ve Growth	No Change
Asset Ownership	52.03	0.00	47.97	84.27	0.00	15.73	0.00	0.00	100.00
Bathroom Facility	19.85	0.00	80.15	46.26	6.80	46.94	0.00	10.53	89.47
Fuel for Cooking	10.42	0.00	89.58	12.88	14.11	73.01	0.00	0.00	100.00
Condition of Household	31.67	0.00	68.33	28.87	7.75	63.38	0.00	15.91	84.09
Main Source of Light	29.90	0.00	70.10	62.56	1.54	35.90	0.00	2.94	97.06
Main Source of Water	48.75	0.00	51.25	9.61	0.00	90.39	0.00	4.90	95.10
Female Emp (Main)	23.08	0.00	76.92	16.75	13.88	69.38	0.00	9.09	90.91
Female Emp (Marginal)	10.47	0.00	89.53	7.08	57.52	35.40	0.00	52.28	47.72

Table 3: Shown is the percentage of districts at each level, that have shown a positive movement to a higher level, or a negative movement, or no change from their current level. The table shows the percentage values for all the indicators

4 PATTERNS OF DEVELOPMENT

We show in Figure 2 changes in the dominant type of employment in districts between 2001 and 2011. Only a few districts have actually changed over the decade on this variable.

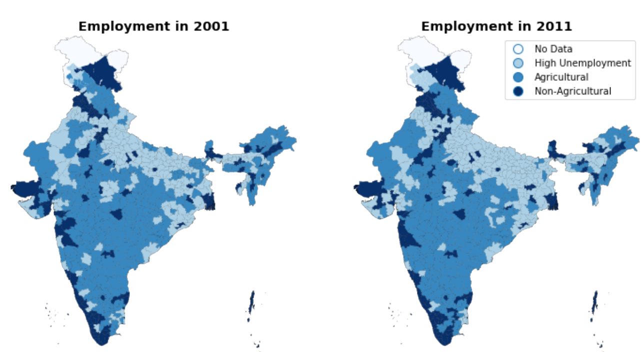

Figure 2: Districts are colour-coded on their dominant type of employment in 2001 and 2011

Only 23 districts which had a high degree of unemployment in 2001 were able to move to a predominantly agricultural or non-agricultural type of employment. Similarly, only 5 districts changed from a predominantly agricultural employment profile to a non-agricultural employment profile. A few districts even moved from a non-agricultural to agricultural profile, possibly indicating that there was little growth in non-agricultural employment in these districts causing the growing number of households there to fall back on their agricultural inheritance instead of being able to find employment in other sectors or even migrate outside the district. Most districts (87%) stayed the same in terms of their employment profile.

Variable	Existing Status	Non Agricultural	Agricultural	High Unemployment	Total
Asset Ownership	Level-1	0.895	0.498	0.472	0.571
	Level-2	0.816	0.714	0.962	
	Total	0.851	0.509	0.534	
Bathroom Facility	Level-1	0.742	0.142	0.172	0.268
	Level-2	0.8	0.171	0.333	
	Total	0.779	0.146	0.213	
Fuel for Cooking	Level-1	0.317	0.076	0.078	0.111
	Level-2	0.688	0	0.093	
	Total	0.421	0.064	0.086	
Main Source of Light	Level-1	0.833	0.315	0.261	0.462
	Level-2	0.741	0.636	0.54	
	Total	0.758	0.513	0.345	
Condition of Household	Level-1	0.714	0.36	0.19	0.3
	Level-2	0.591	0.234	0.144	
	Total	0.621	0.289	0.168	
Main Source of Water	Level-1	0.212	0.588	0.511	0.257
	Level-2	0.333	0.072	0.066	
	Total	0.263	0.325	0.189	

Table 4: Shown is the probability of positive change in indicators for districts having different types of employment. This is disaggregated further into the probability of positive change based on the existing status of the districts. The last column shows the overall probability of positive change

Table 4 shows the probability of positive change in an indicator given the type of employment of a district, ie. P(positive change | type of employment). A movement from level-1 to level-2 or level-3, or a movement from level-2 to level-3, is considered as a positive change. All other changes are considered as non-positive changes. For example, considering the indicator for asset ownership, the last column shows P(positive change) for asset ownership = 0.571, calculated across all districts. P(positive change | type of employment) in asset ownership is shown in the last row as 0.851, 0.509, and 0.534, for the changes in non-agricultural, agricultural, and high unemployment districts respectively. A further dis-aggregation is shown for P(positive change | type of employment, current status) considering the current status of a district of being at level-1 or level-2 in asset ownership. The P(positive change) for all other variables was calculated in a similar manner. We analyze this table carefully to determine several broad patterns of development, as explained next. Note that level-3 districts are not shown in the table because they are already at the most developed level, and very few negative movements were recorded from this level.

We also check whether non-agricultural districts are more likely to be at level-3 for various socio-economic indicators. Table 5 shows the % of districts at level-3 (as of 2011) for each variable. We see that for four out of six variables, non-agricultural districts have a higher share of level-3 districts.

Variable	Non Agricultural	Agricultural	High Unemployment
Asset Ownership	64.38	15.75	19.86
Bathroom Facility	64.07	17.37	18.56
Fuel for Cooking	72.66	11.72	15.63
Condition of Household	48.52	39.05	12.43
Main Source of Light	37.69	43.30	19.00
Main Source of Water	34.47	46.38	19.15

Table 5: Shown in the percentage of districts at level-3 (as of 2011) for the three types of districts

4.1 Hypothesis 1

Non-agricultural districts see the greatest improvement in all indicators.

Observing Table 4, we can see that out of the six indicators, non-agricultural districts see the highest probability for positive change in five of them. Further dis-aggregating based on the current status of the districts, shows that this hypothesis is true irrespective of the current status. Only for the main source of water, do we see that agricultural districts have a greater tendency for positive change, which is also predominantly due to level-1 agricultural districts having progressed rapidly. There were 53 such districts, many of them from central India, indicating that the hypothesis was invalidated in only a few districts which potentially saw special policy attention being given on them. We carried out one-sided Z-tests to test the hypothesis, by taking pairs of different employment types. For example, for non-agricultural and agricultural types of employment, we created two samples of districts and calculated the probability p_1 for a positive change in an indicator for non-agricultural districts, similarly p_2 as the probability for a positive change in the indicator for agricultural districts. We then tested the null hypothesis of $p_1 = p_2$, against the alternate hypothesis of $p_1 > p_2$. These tests were repeated for every pair of employment type, and for each indicator. For the test between non-agricultural and agricultural districts, the Z-score and p-value for main source of water were -0.52 and 0.302 respectively and therefore the null hypothesis could not be rejected. The z-scores for the other attributes ranged between 10-19 (much larger than the value of 1.96 considered for tests at a 95% confidence) and p-values were less than 10^{-24}, strongly rejecting the null hypothesis.Detailed calculations for all the tests are shown in the supplementary material. [16].

This observation is not surprising considering that as per the 2018 India Wage Report by the International Labour Organization (ILO), there were large wage disparities between workers based on occupation, literacy, and gender [31]. The average daily wages of regular workers in the primary sector (agriculture and allied activities) was only ₹ 192, while the wages in the secondary and tertiary sectors were ₹ 357 and ₹ 424 respectively [31]. This substantial wage disparity would provide higher disposable income to households engaged in non-agricultural activities, to improve their indicators faster.

Combined with the observation in Table 5 that non-agricultural districts also tend to be at a higher level on various indicators than agricultural and high-unemployment districts, this shows a trend towards increasing inequality. The better-off districts are moving more rapidly towards an improved life than less well-off districts, highlighting that less well-off districts may require policy attention so that they are not left behind for long. This phenomenon of rising inequality has been explained by the economist Simon Kuznets [15, 25] who states that as an economy develops, market forces first increase and then decrease economic inequality, and is depicted by the inverted U-shaped Kuznets curve. This happens because when an economy moves from agricultural to non-agricultural employment, an influx of cheap rural labour leads to diminished wages and rising inequality. As the economy industrializes further to absorb surplus labour, and aided also by social welfare mechanisms, the inequality is expected to decrease. Based on our data analysis, during 2001 to 2011 India seems to have been on the rising part of the Kuznets curve of increasing inequality. As we show later, this inequality is now observable across districts because industrialization has not spread in a spatially equitable manner in India and has remained concentrated in the same geographical regions.

4.2 Hypothesis 2

Households prefer to invest in assets first, followed by investment in other indicators which they can influence through their own choices.

Out of the six indicators, four of them (asset ownership, bathroom facility, fuel for cooking, and condition of household) are likely to be governed by choices made by households of where to invest their disposable income: Should we invest in purchasing assets, or improve sanitation facilities at our home, or use a better fuel for cooking? We call these discretionary variables because they seem to be more about household preferences than limited by government investments in social infrastructure. Although the choice of bathroom facility and fuel for cooking can be influenced by government assistance, such as by providing sewer connections for bathrooms or distributing LPG for cooking fuel, but irrespective of this government support an individual household can still upgrade itself if needed. For example, pit latrines, covered slabs, or septic tanks can be used instead of piped sewers, or instead of an LPG connection healthier alternatives for fuel for cooking can be used such as coal, or kerosene.

Figure 3: Change in the levels of districts between 2001 and 2011, for the discretionary variables of asset ownership, bathroom facility, and fuel for cooking.

To find out the relative investment preferences among these four variables, we again review Table 4. We can clearly see that asset ownership seems to be changing the fastest irrespective of the type of the district. This remains consistent even when dis-aggregated based on the current status of the variables. People seem to invest more readily for assets than for other amenities that would lead to a healthier lifestyle for the household. We visualize this in Figure 3 which shows on a map how asset ownership shows more positive change (most amount of green) than bathroom facilities and fuel for cooking (least amount of green).

Similar to the previous hypothesis, we conducted one-sided Z-tests to compare the probability of positive change in asset ownership with the probability of positive change in other discretionary variables. The Z-scores and p-values as shown in Table 6 strongly reject the null hypothesis that these change probabilities are the same.

Bathroom Facility		Fuel for Cooking		Condition of Household	
Z-score	p-value	Z-score	p-value	Z-score	p-value
9.114	3.96E-20	15.415	6.46E-54	8.872	3.58E-19

Table 6: Statistical test for Hypothesis-2 : Z-scores and p-values for a comparison between the probability of positive change in asset ownership with the probability of positive change of other discretionary variables

This choice can be influenced by many factors. Consumer behaviour research shows that personal, interpersonal, and cultural effects can shape asset acquisition and usage, and in particular assets are often seen as status symbols [27]. Alternately, assets such as mobile phones could be economically useful as well, as reported in the context of fishermen and wholesalers in South India [21]. Further, the amount of disposable income may also pay a role: A cheap mobile phone would cost much less than installing a simple water filter, for example. This raises the need for further studies to understand the reasons behind their preferences on such expenditure.

Another pattern which emerges from Table 4 is that districts which are already at level-2 have better chances of improving their level. Consider asset ownership for which the P(positive change | agricultural districts) at level-2 (0.714) is higher than P(positive change | agricultural districts) at level-1 (0.498), and similarly for bathroom facility and other indicators where in general level-2 districts have a higher probability to improve than level-1 districts. This again points towards growing inequality, which augmented with the observation that households prefer to invest in assets over other amenities that might be more important for a healthier life, suggests that suitable policies should be created to not leave the less well-off districts behind and to further nudge households towards making more appropriate investments [38].

4.3 Hypothesis 3

Government has prioritized electrification and lighting over other indicators that depend upon government support.

In the previous hypothesis we considered indicators which can be upgraded based on a household's discretion. We next consider variables which require significant government support, and call them social infrastructure variables. This includes the main source of light, where electrification is predominantly provided by government utility companies, and the main source of water which requires tap water distribution networks or handpumps and wells funded by the government.

Examining P(positive change | type of employment) from Table 4, we observe that both these indicators have improved considerably over the years, demonstrating government support. However, between the two, the main source of light has improved much more, with non-agricultural districts showing the strongest change, but not any significant differences based on the current status of a district. This indicates a consistent effort in infrastructure provisioning by the government irrespective of the current status of a district, but with a bias towards electrification over drinking water provisioning. Figure 4 also shows this visually, about the extent of change in development levels for the main source of light and the main source of water. The one-sided Z-tests gave a Z-Score of 5.97 with a p-value of $1.14 * 10^{-9}$ rejecting the null hypothesis that there is no difference in change between indicators for the main source of light and main source of water.

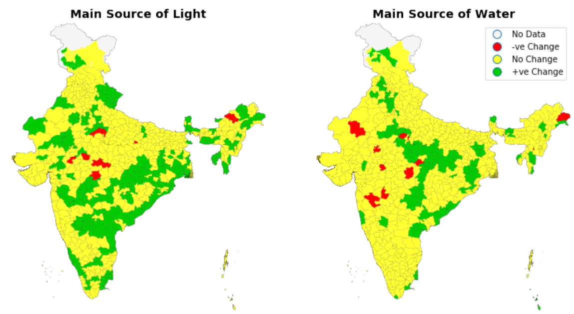

Figure 4: Change in levels of districts between 2001 and 2011, for the social infrastructure variables of main source of light and the main source of water

Historical reports indeed point towards this gap. The introduction of the Electricity Act of 2003 encouraged participation of the private sector in electricity production, and the Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY) launched in 2005 was aimed at creating rural electrification infrastructure so as to electrify all villages and give electricity connections free of charge to families living below the poverty line. The evaluation gave a 93.3% success rate to RGGVY in meeting its targets [11], and access to electricity rose from 59 % of the population in 2000 to 74 % in 2010 [33]. In contrast, the government also launched the Bharat Nirman Program in 2005 with an emphasis on providing drinking water especially to habitations affected by poor water quality, but achieved limited goals with six states reporting less than 50% achievement against targets, and an average achievement of 80.4% [30].

4.4 Hypothesis 4

Discretionary spending by households is closely related to literacy and formal employment and does not seem to be affected by social infrastructure provisioning by the government.

We next check what factors may affect the discretionary spending by households, or rather since we cannot make any causality claims, we check which factors may be stronger predictors of discretionary spending. We check for four factors: literacy, formal employment, current status, and government support for social infrastructure. Variables literacy and formal employment were also discretized into 3 levels of development.Since our variables are discretized into multiple ordinal levels, instead of studying a correlation between pairs of different variables we choose to examine the mutual information between the pairs to form an assessment of the strength of relationship between them.

Literacy	Asset Ownership					
	Non Agricultural		Agricultural		High Unemployment	
	No Change	+ve change	No Change	+ve change	No Change	+ve change
Level-1	0.0101	0.0152	0.1922	0.1417	0.1315	0.1248
Level-2	0.0304	0.0641	0.027	0.0641	0.0287	0.0472
Level-3	0.0388	0.0455	0.0017	0.0236	0.0017	0.0118

Table 7: Shown is the relationship between change in levels of asset ownership with literacy, sliced by the type of employment of the district. This is an example to study the relationship between changes in discretionary variables with factors such as literacy that might predict the change

We first create tables by calculating probabilities of +ve change/no change, such as the one shown for the relationship between literacy and change in asset ownership, in Table 7 for the three types of districts. The sum of the probabilities in each table is equal to 1. We then calculate the mutual information between the two variables:

$$I(X;Y) = \sum_{y \epsilon Y} \sum_{x \epsilon X} p(x,y) \log(\frac{p(x,y)}{p(x)p(y)})$$

where $p(x,y)$ is the joint probability function of x and y, and $p(x)$ and $p(y)$ are the marginal probabilities of x and y respectively. In Table 7, x ranges over the six classes of asset ownership and y ranges over the three levels of literacy. The same method is used to calculate mutual information between each of the four factors of interest in this hypothesis (literacy, formal employment, current status, and government support for social infrastructure), and change in each of the four discretionary variables (asset ownership, bathroom facility, fuel for cooking, and condition of household). These sixteen tables are given in the supplementary information [16].

Figure 5 shows the mutual information calculated on these tables. We can see that literacy, formal employment, and the current status are all more predictive of change in discretionary variables as compared to government support on the main source of lighting and main source of water.

This is not altogether surprising because literacy and formal employment go hand in hand with each other, and the wages in formal employment are substantially higher than wages in the unorganized sector, which puts more disposable income in the hands of people that can be used for discretionary expenditure. The 2018 India Wage report by ILO [31] reveals that the average daily wages of workers in formal employment (organized sector) are ₹ 513, while wages in the unorganized sector are ₹ 166. Similarly, wages for men with the highest level of education as graduates are ₹ 735, while for casual urban workers with the same level of education the wages are ₹ 219.

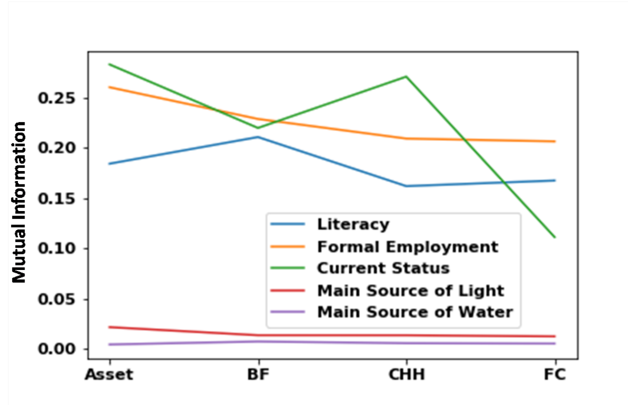

Figure 5: Mutual information is derived between change in discretionary variables (x-axis) and factors that might predict the change. The factors shown here are literacy, formal employment, current status of the discretionary variable, and change in the social infrastructure variables of the main source of light and the main source of water

Similar gaps exist in wages for different levels of education, and also between educated women in formal and unorganized sector employment. A deeper analysis indeed validates that districts with high formal employment in general are more likely to change positively in discretionary indicators. A similar pattern is observed for literacy. Formal employment and literacy are also correlated with each other, with a Pearson correlation coefficient of 0.61. This validates the observations that formal employment and literacy go hand in hand, and formal employment leads to more disposable income that can be spent on discretionary variables.

An interesting insight however is that government spending towards social infrastructure provisioning does not seem to be related to discretionary spending, emphasizing even more on the need to increase disposable income to bring about positive changes in living conditions for people.

4.5 Hypothesis 5

Districts with more manufacturing and services industries end up developing faster. However, the presence of these industries has not spread geographically and has remained spatially concentrated in the same regions over the years.

We observed earlier that non-agricultural districts saw the fastest improvement in all indicators, other than the main source of water. Non-agricultural employment can be further divided into several industries such as manufacturing, retail, real estate, banking, etc. We aggregated these industry sectors into four categories shown in Table 8, and as before we then did a k-means clustering based on the percentage of population employed in each of these categories.

(a)	**Business activity**: Includes the sectors of retail, wholesale, transportation , storage, hotel and restaurant etc.
(b)	**Services**: Includes services sectors like electricity, gas, water supply, defence forces, administration, social security, education, health and social work etc
(c)	**Construction and mining** sector
(d)	**Manufacturing** sector

Table 8: Industry sectors aggregated into four broad categories

We were thus able to label each district in terms of the dominant nature of employment in the district. A good clustering was obtained for k = 4, and a boxplot for these district-types is shown in Figure 6, labeled for a high presence of manufacturing industries (Type-4), services industries (Type-3), a moderate industry presence (Type-2), and low industry presence (Type-1).

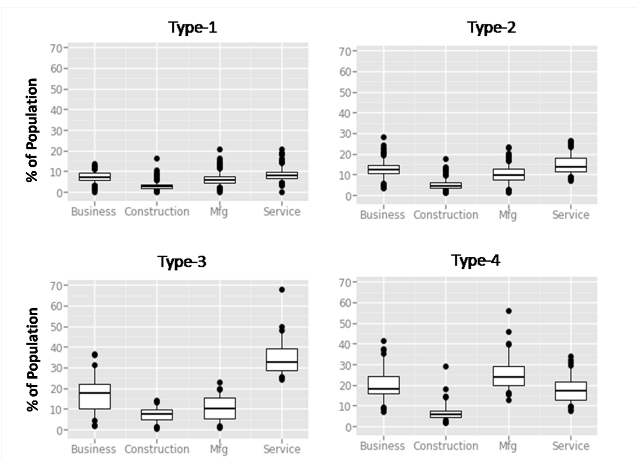

Figure 6: Shown is the distribution of population employed in four industrial categories (business, construction, manufacturing, and services), across four district clusters. The clusters are used to label districts in terms of their dominant type of industrial employment

Table 9 shows the change in various discretionary and social infrastructure variables, with respect to the type of industries in the district. We can make out a clear pattern that type-3 and type-4 districts (having services and manufacturing industries) are more likely to grow faster in all the variables except the main source of water.

These industries also seem to favour formal employment, as seen with a correlation coefficient of 0.77. Here the Z-test was carried out with p_1 representing the probability of a positive change in type-3 and type-4 districts, and p_2 representing that for type-1 and type-2 districts, for each indicator. For main source of water the Z-scores and p-values were 0.34 and 0.362 respectively and therefore the null hypothesis could not be rejected. Z-scores for the other variables ranged between 8-18 (higher than the threshold of 1.96 considered for tests at a 95% confidence) with p-values less than 10^{-16}, strongly rejecting the null hypothesis of $p_1 = p_2$. Results are reported in detail in the supplementary material [16].

Variable	Existing Status	Type1	Type2	Type 3	Type 4	Total
Asset Ownership	Level-1	0.385	0.678	0.714	0.867	0.571
	Level-2	0.667	0.872	0.875	0.783	
	Total	0.388	0.726	0.784	0.83	
Bathroom Facility	Level-1	0.066	0.369	0.4	0.517	0.269
	Level-2	0.121	0.383	0.788	0.714	
	Total	0.072	0.374	0.737	0.6	
Fuel for Cooking	Level-1	0.032	0.132	0.19	0.478	0.111
	Level-2	0	0.236	0.5	0.389	
	Total	0.022	0.165	0.217	0.439	
Main Source of Light	Level-1	0.164	0.574	1	0.8	0.462
	Level-2	0.586	0.662	0.692	0.667	
	Total	0.343	0.622	0.714	0.696	
Main Source of Water	Level-1	0.592	0.5	0.25	0.267	0.25
	Level-2	0.05	0.141	0.333	0.238	
	Total	0.231	0.305	0.261	0.25	
Condition of Household	Level-1	0.224	0.375	0.75	0.636	0.3
	Level-2	0.107	0.361	0.607	0.447	
	Total	0.169	0.366	0.65	0.49	

Table 9: Shown is the probability of positive change in indicators for districts having different types of industrial employment. This is also disaggregated into the probability of positive change based on the existing status of districts. The overall probability of positive change for an indicator is given in the last column

Put together with the earlier hypotheses, this shows that discretionary variables tend to improve with improvements in formal employment in the manufacturing and services sectors by generating disposable income. These sectors are likely to employ more literate people and hence we see a strong correlation between literacy and formal employment. Among discretionary variables people seem to choose to invest in assets before other essential amenities, possibly pointing to the need for policies to nudge behavior towards making appropriate household investments. Finally, all of these variables seem to improve faster in level-2 districts than level-1 districts, which raises a concern about growing inequality. Given the importance therefore of formal employment in the manufacturing and services sectors in the overall development process, we go on to investigate patterns in the growth of these sectors over the years between 2001 to 2011.

Figure 7 plots the dominant industrial type in each district, for the years of 2001 and 2011. There has clearly not been much change over an entire decade. This shows that the manufacturing and services sectors have not expanded to other geographies, an observation also made in other studies [24, 26].

A few regions in Maharashtra (close to Mumbai) and Tamil Nadu (garments and textile industry) have seen a spatial expansion of manufacturing and services industries, likely due to a spillover to neighbouring districts as a consequence

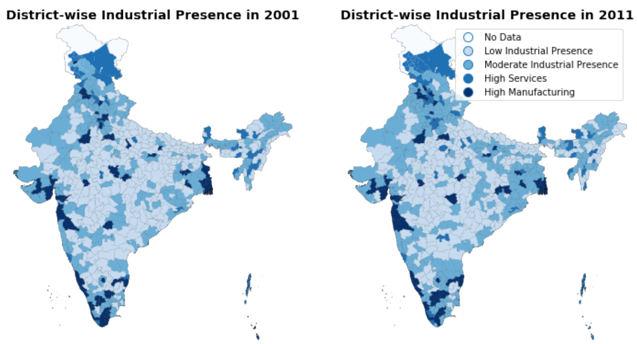

Figure 7: Districts are colour coded based on their type of industrial employment in 2001 and 2011

of growth in the industries. Some new hubs seem to have emerged in Maharashtra (Nagpur) and Orissa (Sundargarh), both of which are known to have strong local industries. An increase is also seen from a low industrial presence to a medium presence in large areas of Tamil Nadu, Orissa, Eastern Uttar Pradesh, and Bihar. Interestingly regions in West Bengal (Murshidabad) and Madhya Pradesh (Sagar) have actually seen a decrease. Overall however, this observation points towards the need for policies to create more widespread non-agricultural employment. This does not however answer questions about the consequence of this spatially concentrated growth in non-agricultural employment. Other studies show that it has led to an increase in rural-urban migration, which tends to be exploitative, and diminishes the ability to distribute economic growth equitably across the country [12].

4.6 Hypothesis 6

Female participation in the workforce has decreased, primarily with a reduction in marginal employment that has not been compensated with an equivalent increase in female main employment.

Earlier in Table 3 we observed a sharp fall for female employment as marginal workers (less than six months of employment in a year). This could be either due to an increase in formalization that saw more women moving from marginal workers to main workers (more than six months of employment in a year), or an overall decrease in female employment itself.

Table 10 shows a confusion matrix for the change in levels for main and marginal female workers between 2001 and 2011. We can see that there are 210 districts which saw a fall in female marginal employment but with no change in main employment, and an additional 28 districts which saw a fall even in the main employment of women. Only 83 districts have actually shown an increase in their levels

Female Employment main	Female Employment marginal		
	+ve change	-ve change	No change
+ve change	7	33	43
-ve change	1	28	16
No change	9	210	246

Table 10: Shown is the number of districts that have changed positively, negatively, or not changed in their level for the variable of female main employment, against changes in their level for the variable of female marginal employment

for female main employment. This seems to indicate that although there is a positive movement from marginal to main employment in some districts, but overwhelmingly more often women are actually falling out from the workforce, especially in marginal employment. The one-sided Z-tests corresponding to this hypothesis for falling female marginal employment were performed in a different manner. For each level, we calculated p_1 as the probability for the female marginal employment to be at that level in 2011, and similarly p_2 as the probability for the marginal employment to be at that level in 2001. The Z-scores and p-values for these tests are shown in Table 11. Since the Z-score for districts at level-3 in female marginal employment is highly negative, while it is highly positive for level-1 and level-2 districts, it shows that indeed districts have moved from level-3 (high marginal employment) to lower levels. However, when the same test is performed for female main employment, no conclusive observations can be made other a positive change in female main employment in level-3 districts only.

Female Employment	**Marginal**		**Main**	
	Z-score	**p-value**	**Z-score**	**p-value**
Level-1	7.562	1.98E-14	-0.859	0.195
Level-2	5.671	7.07E-09	-0.427	0.334
Level-3	-11.441	1.31E-30	1.313	0.094

Table 11: Statistical tests corresponding to Hypothesis-6: Shown are the Z-scores and p-values for changes between 2001 and 2011 in the levels of female marginal and main employment

Several studies try to explain the falling out of women from the workforce [36, 37] and cite reasons such as poor working conditions and wage disparities between men and women, which demotivate women to take up work, especially as women are getting more educated. Some models of development that relate economic growth with gender equality suggest that this is actually expected [17, 28]. As household income increases at the same time as the economy moves from agricultural to non-agricultural employment, women who earlier were involved in agricultural work begin to move out from the workforce. Eventually as education and formal employment increases, women begin to enter the workforce again, leading to a U-shaped function for female employment. According to our data analysis, India seems to have been on the falling part of the curve during 2001 to 2011. Upon further observing these changes in female employment patterns for different types of districts, we find that agricultural and non-agricultural districts are not very different from each other. Both have seen a large fraction of districts (37% and 41% respective) reduce in female marginal employment, and a small fraction of districts (16% and 10% respective) increase in female main employment. This seems counter-intuitive because non-agricultural districts should expect to see a stronger movement to main employment than agricultural districts; the phenomenon therefore needs to be monitored and further studies should be conducted to check whether the transition from marginal to main employment for women is happening in a robust manner as predicted by the U-shaped hypothesis.

5 DISCUSSION AND CONCLUSIONS

We were able to make several interesting observations from analysis of the census data between 2001 and 2011. We saw a clear link between non-agricultural employment in the manufacturing and services sectors, with literacy and household spending for assets. We saw that these industrial sectors have not expanded to other geographies but have remained spatially concentrated in their pre-existing locations. We saw that households prefer to spend on assets before they spend on other amenities such as the fuel for cooking, bathroom facilities, and physical condition of their households. We saw that the government provides support on social infrastructure such as electrification and drinking water, but puts more preference on electrification. We saw that this government support however does not seem to influence discretionary spending by households. We saw that female participation in the workforce has reduced, and we are able to associate it with a reduction in marginal employment that has not been compensated with an increase in main employment as yet, irrespective of the dominant type of employment in a district. The underlying mechanisms of why these patterns emerge and what consequences they lead to, need to be investigated further. Many studies have investigated these patterns but several questions remain unanswered. This leads us to suggest that data analysis of sources like census data, including also much work happening with the use of big-data sources such as satellite data, commodity prices, and cellphone call records, can certainly reveal interesting observations. However, these observations need to be investigated further through ethnographic and survey studies to understand the dynamics that might be leading to the observations and resulting from them. We feel that newspaper reports, social media, and other participatory media networks where public conversations take place between people, may give hints about these underlying mechanisms. So far, studies have used social media and mass media data to build indicators such as for unemployment [2], economic uncertainty [5], and trade and retail [9]. However, studies have not been done to explain observations that are being noticed through a purely statistical analysis of different data sources. We feel that this can be a rich area for future research, where observations made through especially big-data sources about interesting anomalies and correlations between variables, can trigger specific analysis in qualitative data from media networks that publish information about the lives of people. Combining the mechanisms highlighted through such qualitative data analysis, with the observations made through quantitative big-data analysis, can help build a comprehensive district development model that can be explicitly evaluated on values deemed to be more important by policy makers.

We have taken a small step in this direction of working with a social enterprise, Gram Vaani, that runs voice-based participatory media networks in both, rural areas of India and a few urban centers, for low-income communities [29]. Through ordinary mobile phones, people make calls into these platforms to listen to audio messages contributed by other users, and also leave their own messages. These messages are reviewed by a team of moderators and then published back on to the platform. Active discussions happen about policies, current affairs, government schemes, agriculture, employment, etc, and are often even seeded by the Gram Vaani team to solicit views and experiences of the community members on specific topics. A discussion campaign was launched recently called *Shram ka Samman*, which translates to *Dignity of Labour*, in both rural areas (which are sources of migration) and urban areas (having significant industrial presence, which see an inflow of rural migrants). A qualitative report published on the first phase of the campaign [40] is able to highlight several dynamics such as the exploitative work-relations for rural migrants coming for work to urban areas, and the lack of local employment opportunities which leads to migration. We feel that a rich area for future work lies ahead in automatically analyzing conversations happening on such participatory media platforms, and linking them with observations made from different big-data sources, to mutually complement one another. District development models thus built can reveal the underlying dynamics unfolding in the lives of people, and provide a means to evaluate the models based on different value-systems.

6 ACKNOWLEDGMENTS

We are very thankful to the anonymous reviewers and our shepherd, Afsaneh Doryab, for their feedback to improve the paper. We also thank the IIT Delhi HPC facility for computational resources used for the research.

REFERENCES

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. 1993. Mining association rules between sets of items in large databases. In *Acm sigmod record*, Vol. 22. ACM, 207–216.
[2] Dolan Antenucci, Michael Cafarella, Margaret Levenstein, Christopher Ré, and Matthew D Shapiro. 2014. *Using social media to measure labor market flows.* Technical Report. National Bureau of Economic Research.
[3] Sam Asher, Karan Nagpal, and Paul Novosad. 2017. The Cost of Distance: Geography and Governance in Rural India. (2017).
[4] Carmel Cahill Ashok Gulati. 2018. Resolving the farmer-consumer binary. https://bit.ly/2Nz8bS5.
[5] Scott R Baker, Nicholas Bloom, and Steven J Davis. 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131, 4 (2016), 1593–1636.
[6] Joshua Blumenstock, Gabriel Cadamuro, and Robert On. 2015. Predicting poverty and wealth from mobile phone metadata. *Science* 350, 6264 (2015), 1073–1076.
[7] Joshua E Blumenstock and Niall Keleher. 2015. The Price is Right?: Statistical evaluation of a crowd-sourced market information system in Liberia. In *Proceedings of the 2015 Annual Symposium on Computing for Development.* ACM, 117–125.
[8] C Chandramouli and Registrar General. 2011. Census of India, 2011. *Provisional Population Totals. New Delhi: Government of India* (2011).
[9] Hyunyoung Choi and Hal Varian. 2012. Predicting the present with Google Trends. *Economic Record* 88 (2012), 2–9.
[10] Planning Commission et al. 2011. India human development report 2011: Towards social inclusion. *Institute of Applied Manpower Research, Government of India Google Scholar* (2011).
[11] Planning Commission et al. 2014. Evaluation Report on Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY). (2014).
[12] Arjan De Haan. 2011. *Inclusive Growth?: Labour Migration and Poverty in India.* International Institute of Social Studies.
[13] Gasper Des. 2004. The ethics of development. From economism to human development. *Edinburgh Studies in World Ethics, Edinburgh University Press, Edinburgh* (2004).
[14] Priya Deshingkar, Rajiv Khandelwal, and John Farrington. 2008. *Support for migrant workers: The missing link in India's development.* ODI London.
[15] Amit Basole et al. 2018. State of Working India 2018. https://bit.ly/2Fb4TC6.
[16] Dibyajyoti et al. 2019. Supplementary Material : Towards Building a District Development Model Using Census Data. http://bit.ly/31EADcf.
[17] Claudia Goldin. 1994. *The U-shaped female labor force function in economic development and economic history.* Technical Report. National Bureau of Economic Research.
[18] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028.
[19] FarmGuide India. 2017. Small and Marginal Land Holdings, Farmer Challenges. https://bit.ly/2W11ag5.
[20] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.
[21] Robert Jensen. 2007. The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *The quarterly journal of economics* 122, 3 (2007), 879–924.
[22] Kunal Keshri and Ram B Bhagat. 2013. Socioeconomic determinants of temporary labour migration in India: A regional analysis. *Asian Population Studies* 9, 2 (2013), 175–195.
[23] Kalpana Kochhar, Utsav Kumar, Raghuram Rajan, Arvind Subramanian, and Ioannis Tokatlidis. 2006. India's pattern of development: What happened, what follows? *Journal of Monetary Economics* 53, 5 (2006), 981–1019.
[24] Anirudh Krishna. 2017. *The broken ladder: The paradox and potential of India's one-billion.* Cambridge University Press.
[25] Simon Kuznets and John Thomas Murphy. 1966. *Modern economic growth: Rate, structure, and spread.* Vol. 2. Yale University Press New Haven.
[26] Somik Vinay Lall and Sanjoy Chakravorty. 2005. Industrial location and spatial inequality: Theory and evidence from India. *Review of Development Economics* 9, 1 (2005), 47–68.
[27] Peter Ling, Steven D'Alessandro, and Hume Winzar. 2015. *Consumer Behaviour in Action.* Oxford University Press Oxford.
[28] Kristin Mammen and Christina Paxson. 2000. Women's work and economic development. *Journal of economic perspectives* 14, 4 (2000), 141–164.
[29] Aparna Moitra, Vishnupriya Das, Gram Vaani, Archna Kumar, and Aaditeshwar Seth. 2016. Design Lessons from Creating a Mobile-based Community Media Platform in Rural India. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development.* ACM, 14.
[30] Ministry of Drinking Water and Government of India Sanitation. 2011. Report of the Working Group on Rural Domestic Water and Sanitation. https://bit.ly/2JyR7g8.
[31] International Labour Organistaion. 2018. India Wage Report, Wage policies for decent work and inclusive growth. https://bit.ly/2wjhQ80.
[32] Amitendu Palit. 2013. Indian Economy, Amartya Sen and Jagdish Bhagwati. https://nus.edu/2HwSGti.
[33] Sheoli Pargal and Sudeshna Ghosh Banerjee. 2014. *More power to India: The challenge of electricity distribution.* The World Bank.
[34] Divya Varma Rameez Abbas. 2014. Internal Labor Migration in India Raises Integration Challenges for Migrants. https://bit.ly/2HlEp3h.
[35] Hans Rosling. [n. d.]. With Ola Rosling and Anna Rosling Rönnlund. 2018. *Factfulness: Ten Reasons We're Wrong About the World And Why Things Are Better Than You Think* ([n. d.]).
[36] Prachi Salve. 2019. Why Rural Women Are Falling Out Of India's Workforce At Faster Rates Than Urban Women. https://bit.ly/2FeZc5d.
[37] Sunita Sanghi, A Srija, and Shirke Shrinivas Vijay. 2015. Decline in rural female labour force participation in India: A relook into the causes. *Vikalpa* 40, 3 (2015), 255–268.
[38] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth and happiness.* Penguin.
[39] Jayan Jose Thomas, MP Jayesh, et al. 2016. Changes in India's Rural Labour Market in the 2000s: Evidence from the Census of India and the National Sample Survey. *Journal* 6, 1 (2016), 81–115.
[40] Vani Viswanathan. 2019. Shram ka Samman: On Migrants Hopes, Rights and Demands. http://gramvaani.org/?p=3244.
[41] Tricia Wang. 2016. Why Big Data Needs Thick Data. https://bit.ly/23E9qlv.