# Policy for Big Data: An Investigation Using Land Records

**Sachin Garg***

***PhD Candidate, Schar School of Policy & Government***
***George Mason University***
***E-mail: sgarg6@gmu.edu***

## Abstract

In an increasingly digitised world, Information and Communication Technologies (ICTs) are playing a key role in human development. A consequence of this increased digitisation is the exponential increase in the being created. Data from multiple sources can be "mashed up" and processed using advanced analytics to provide deep and detailed insights on the world around us. However, there are many challenges in the use of "Big Data for Policy", a key one being around the "Policy for Big Data". An often understated aspect is the fact that much of this "big data" (especially in the emerging economies) is owned by the private sector that has the necessary wherewithal to collect and analyse it.

Nevertheless, during the process of governing, the public sector creates "administrative data", which is a unique resource. However, even creating administrative data in a form that can be linked together to create "big data" is fraught with challenges. The aim of this research is identify these challenges so that policy makers can attempt to mitigate them.

Using the domain of land records, I have created a novel dataset on the proliferation of a land computerisation project in the states of India, seeking to identify the key factors behind the uneven uptake of this project in the states. I hypothesise that the diffusion of policies that seek to create digital records depend on three main factors-—the relative level of socio-economic development, the amount of support a policy would have amongst the populace and the complexity of policy adoption. The research finds significant evidence for all the three hypotheses. ***Keywords—*** Big Data; administrative data;land records; India; policy diffusion; DILRMP; NLRMP.

## 1 Introduction

In an increasingly digitised world, Information and Communication Technologies (ICTs) are becoming key to development (World Bank, 2016). An outcome of this increased digitisation has been the "data tsunami" (Decker, 2014) — an exponential increase in the amount of data being created. Thanks to the rapid advances in computing, we now have the capability to crunch ever increasing amounts of data to better understand the world around us. Moreover, data from multiple sources can be "mashed" together, allowing us to get more detailed pictures of our world than was possible earlier. It is this increased amount of data, coupled with analytics that goes by the name of "Big Data" (Ward & Barker, 2013). The extant literature has looked at how such "big data" can be leveraged for policy making (Desouza & Jacob, 2014) and human development (UN Global Pulse, 2012). Hilbert

(2013) places the "analytic treatment of data … at the forefront of intelligent decision making". However, there are also many challenges in the use of use of big data for policy making that are related to proprietary algorithms (Lazer, Kennedy, King, & Vespignani, 2014), privacy concerns (Daries et al., 2014; Lane, Stodden, Bender, & Nissenbaum, 2014; Schintler & Kulkarni, 2014) and the fact that much of the "big data" is in the hands of private sector enterprises that possess the necessary wherewithal to collect and analyse this big data (Taylor & Broeders, 2015).

On the other hand, the public sector, routinely collects "administrative data" (census, health surveys, economic and others) that can be digitised and linked together can form a unique source of "Big Data". This "Big Data" can then be used to analyse, design and monitor public policy (Taylor, Cowls, Schroeder, & Meyer, 2014). However, creating this "Big Data" requires explicit policies that allow gathering of the constituent data and its linkage. Thus, the public sector faces the conundrum that we need "Big Data for Policy" while simultaneously also needing a "Policy for Big Data".

One source of administrative data is land ownership and use records, called the "cadastre". Countries are in the process of computerising the cadastre and linking it to other forms of data, especially spatial data, to gain a fuller understanding of the resources they possess and administer them better. However, there are huge variations in how countries (both at the national and sub-national levels) are proceeding in this task. This research attempts to identify some of the factors that lead to such variations at the sub-national level. Using a novel data set (created by the author), I look at how Indian states are implementing a centrally-sponsored program to computerise land records. India's "National Land Records Modernisation Programme" was started in 2008 and supports states in their land records computerisation efforts by providing them technical and financial support.

The rest of this paper is structured as follows. Section 2 provides a very brief introduction to the domain of land records and why I think of the domain as creating "quintessential" Big Data. Section 3 discusses the research questions and hypotheses, while the preliminary results are discussed in section 4. Some ideas for future work are put forth in section 5.

## 2 The Land Records Domain

Land is undoubtedly the most valuable possession of human society and individuals and effective land policies play a key role in human development (Feder & Feeny, 1991; Feder & Nishio, 1998). Land policy can be thought of as a "governmental instrument" that states the "strategy and objectives for the social, economic and environmental use of the land and natural resources of a country", or "land management implements land policy by means of land administration" Törhönen (2004). The main purposes of land administration are (a) regulating land and property development, (b) using and conserving land, (c) gathering revenue from the land, and (d) conflict resolution, which can be categorised as —juridical, regulatory, fiscal and information management functions (Dale & Mclaughlin, 1999). These functions are distributed amongst specialized agencies that survey and map the land, register it and appraise it. According to Törhönen (2004), land administration is key to proper land management as it "enhances legality and provides information about the land" by maintaining the cadastre. A lack of clarity in land records and disputes can result in productivity losses[1].

The cadastre can be thought of as a "uniquely delimited tract of land within which a coherent set

[1] The McKinsey Global Institute estimated that land market distortions in India accounted for up to 1.3% of lost growth annually (MGI, 2001)

of definable property rights is recognized" (Dale & Mclaughlin, 1999). Navratil and Frank (2004) hold cadastres key to (a) the guarantee of property rights, i.e. ownership, mortgages, encumbrances etc. and (b) property taxation. The cadastre is a general term for official land records, and has two primary parts — the legal register and the fiscal register (Navratil & Frank, 2004; Törhönen, 2004). The legal register contains information relating to the ownership and the bundle of rights associated with of the land parcel, while the fiscal register contains information related to taxation etc. Williamson and Ting (2001) posit that "administration and cadastral systems are a key component of the infrastructure that supports and facilitates the way that society interacts with land to ensure sustainable development".

However, land administration goes beyond ownership and titles, as it also considers "tenure" which is the manner in which the land rights are held (Dale & Mclaughlin, 1999). Tenure is a complex and multifaceted term pulling in connotations from various disciplines (at least) — legal, social, economic and technical Törhönen (2004). Dale and Mclaughlin (1999) define land tenure as a "dense network" of intersecting interests along with associated duties, obligations and powers that form a hierarchy. These interests can be broadly categorised as — *overriding* (like the state's power of eminent domain), *overlapping* (where multiple parties can use the land parcel), *complementary* (shared) or *competing* (contested interest(s) in the same parcel). The concept of tenure allows us to separate ownership from usage — tenure status is the "the mode by which land or property is owned or held", while property rights define "what one is permitted to do with such land or property" (Payne, 2004). This has important implications for many poor people for whom the ability to use the land may be much more important than formally owning it (Dale & Mclaughlin, 1999; Törhönen, 2004). As discussed by Törhönen (2004), the largest category of tenure is a "multi-layer" tenure that is "undefined, multidimensional, ambiguous and elastic" and managing this "multi-layer tenure" is the largest challenge land administrators face.

## 2.1 Multi-Purpose Cadastre

The National Research Council of the US defined a multi-purpose cadastre as a system "designed to overcome the difficulties associated with these more limited approaches by (1) providing in a continuous fashion a comprehensive record of land-related information and (2) presenting this information at the parcel level. The multipurpose cadastre is further conceptualized as a public operationally and administratively integrated land-information system, which supports continuous, readily available, and comprehensive land-related information at the parcel level" (National Research Council, 1980).

The rationale for such a system is multifaceted. For example, the relationship between humans and land is dynamic and changing over time requiring a mechanism that can allow all information about land parcels to be fetched from a single source (Williamson & Ting, 2001). Deininger and Goyal (2012) provide an economic rationale for a "wide[r] availability of reliable information on property rights in land at low cost through land registries" as key to using land as collateral, and thus provide an "ability to transfer assets to more productive users". On the other hand, Törhönen (2004), points out that "security of tenure" is a *real* issue which it is *not* synonymous with statutory recording and identifies the main challenge in land administration and registration as the "*[in]ability to quantify, recognise, record, regulate and manage land tenure, which consists of multiple layers that have volatile formal and informal significances*". However, land tenure is multi-dimensional and as rights can be sliced and diced in multiple ways, identifying the owner(s) of particular right(s) is not simple (Törhönen, 2004). A "Multi-Purpose cadastre" also called a "Land Information System" can provide all stakeholders a multi-dimensional view cutting across the spatial, temporal, financial and legal domains, and thus obviate many of these problems.

## 2.2 Multi-Purpose Cadastre as "Big Data"

The Multi-Purpose Cadastre (MPC) is quintessential "Big Data" in that it integrates multiple sources and structures of data. Although the term "Big Data" defies definition, the defining characteristic of "Big Data" that distinguishes it from "lots of data" (Borne, 2013) is what are called its 3Vs — Volume, Velocity and Variety (Laney, 2001). Apart from the large *Volume* of the data often flowing at *Velocity*, much of it comes from diverse data sources (*Variety*) in multiple formats needing the ability to merge it together often without the presence of explicit identifiers. This merging together of data sources (*data fusion)* and analysis of the fused data is key to getting *actionable intelligence* out of the data. As Hilbert (2013) emphasizes — "[the] crux of the "Big Data paradigm is *actually not* the increasingly large amount of data itself, but its *analysis for intelligent decision-making.*"

The MPC possesses all these features that distinguish "Big Data". As Törhönen (2004) noted, "the land administration and land registration challenge lies in the ability to quantify, recognise, record, regulate and manage land tenure, which consists of multiple layers that have volatile formal and informal significances". hence, Thus, for the MPC to work as intended, it is necessary to fuse multiple data sources to provide the near real-time analysis needed for decision making. The basis of the system is a core Geographical Information System (oftentimes provided by a National Spatial Data Infrastructure (NSDI) (National Research Council, 2007; Williamson & Ting, 2001). This system is then linked with other systems that provide parcel information (possibly from fiscal registers), information about soil and vegetation cover (from the departments of agriculture and forests etc.) and to legal systems to get information about the rights, restrictions and responsibilities. It should be noted that each of these systems could in turn be comprised of multiple sub-systems. Also, by virtue of them being intended for different purposes, the various data sources that form part of the MPC come in a variety of formats, with often no common identifier, as well as having different spatio-temporal dimensions.

Although the case for the Multi-Purpose Cadastre was made by the National Research Council in 1980, it continues to be a work in progress across most jurisdictions. I situate my research into the generation of data in India's attempts to digitise its land records and create an integrated land information system which is discussed next.

## 2.3 National Land Records Modernisation Programme

Computerization of land records is not new as computers were being used for land records management shortly after their commercial availability fifty years ago. This worldwide phenomenon was pioneered by the USA and Australia in the early seventies (Lang, 1981; Maggs, 1973). By the eighties Austria, Ontario (Canada) and India (8 districts) had joined the effort (Ahuja & Habibullah, 2005). In India, systems computerization of current land records systems has largely focused on the "digitization" of existing records, primarily the fiscal cadastre, little attention being paid to the spatial aspects and on-ground position. Thus, information about the actual land parcel that would be pertinent to making land-use decisions has not been readily available. Complicating this is the fact that most land titling in India is "presumptive" as opposed to the "conclusive" or "Torrens" title system[2].

Considering the *prima facie* benefits of the Torrens system, the Government of India started an ambitious plan to move the country's titling system to a Torrens type regime. However, as land administration in India is under the purview of provincial governments, and a Torrens title system has to function as a unified and modern land management system, the government started the "National Land Records Management Programme (NLRMP)" project to provide capacity-building assistance to states in modernising their existing land record systems.

This program is implemented at the district level and provides partial central government funding to the state government(s) for digitising their land records, land survey/re-survey, infrastructure creation and linking the data to GIS. States are responsible for submitting funding proposals under this scheme. However, there is a significant variation in the uptake of the project among the various states as can be seen in Table 1[3]. This variation provides an impetus to understanding what impacts the adoption of policies that attempt to create administrative data which can then be linked together to create "Big Data". This aspect of the research is discussed next.

[2] The challenge with presumptive titling is that there is no "single proof of title" necessitating maintenance of an adequate chain of conveyance deeds to answer any questions regarding the ownership, rights and responsibilities related to a land parcel. However, in the case of "conclusive" (or Torrens) titling, a single register is considered to be *the* source of truth as regards ownership, rights and responsibilities relating to the land.

[3] I am looking at the period between 2008–14 which coincides to the starting of the scheme (2008) to when the government changed at the centre (May, 2014).

# 3 Research Questions and Hypotheses

As discussed earlier, Table 1 shows a significant variation in the uptake of the NLRMP among the states of India. Using theories from the policy diffusion literature, this study attempts to identify the potential reasons behind states not adopting a policy that aims at creating data. A novel dataset has been created by cross-linking data provided by the NLRMP project with other data including national census data , district level land use and other state and national level indicators. The main dependent variable is a dichotomous variable (NLRMP) indicating whether the policy has been adopted. A logistic regression is run on this data. The list of variables is provided in Table 2.

The main question(s) that I am researching are (a) what are the key factors that determine whether a state will be part of the NLRMP?, and (b) what district level factors determine if it is chosen for implementation?. The policy literature has identified both exogenous and endogenous factors that impact policy adoption at the sub-national level. Some of these factors include resource availability (evidenced by measures like GDP), politics, policy specific factors, geographical and ideological distance between states (Karch, 2007; Makse & Volden, 2011; Nicholson-Crotty & Carley, 2016; Sugiyama, 2008). I am considering three main factors — economic, political and administrative capacity. The rationale for this follows.

Policies will only be adopted if a particular jurisdiction has the economic wherewithal to pay for its adoption, therefore adoption is a function of the jurisdiction's development. Further, politicians will only back policies that impact the bulk of their constituents positively. Another factor would be the ability of the administrative bureaucracy to actually implement the policy in question, that is have the "administrative capacity" to deliver, which is dependent on implementation complexity.

### 3.1 Hypothesis I. Resource Availability

Adoption of a policy requires resource allocation, hence, only a jurisdiction with sufficient resources will be able to commit to it. In a policy envisaging states to match central contribution, this is even more so. A similar consideration exists in the case of districts. In the case of an e-governance implementation policy (like the current one), this would be even more marked as a less developed district would not have enough consumers to justify the policy push. The above lead to the following hypothesis.

**H1: The adoption of a development policy imperative is directly proportional to the extent of development.** I test this hypothesis at both the state and district levels. At the state level, the State's Development index is a predictor, while at the district level the District Development Indicator is used as the predictor, while controlling for state level effects.

### 3.2 Hypothesis II. Constituency Effects

In a democracy, policy makers will adopt and implement those policies that have a higher chance of being popular with the electorate. In the case of the land records policy, the impacted areas are largely rural and the impacted constituents are those dependent on agriculture—cultivators and agricultural labourers. I capture this particular construct using two variables: (a) the ratio of the district's rural area normalised to the entire state's rural area (AREA_LQ_RURAL), and (b) the ratio of the district's proportion of total agricultural workforce to the state's proportion of agricultural workforce (TWFRAGRI_LQ_TOT). However, these two variables are highly correlated (Table 4) and their essence is captured by creating a new variable (RURALNESS_IDX) using Principal Component Analysis (Table 5. The following is hypothesised.

**Controlling for development, a more rural jurisdiction will adopt the policy** This hypothesis is tested at the district levels, controlling for both district and state level development indicators.

### 3.3 Hypothesis III. Administrative Capacity

Administrative Capacity plays a key role in policy implementation. Nicholson-Crotty and Carley (2016) have theorised that there are few takers for complex policies that require a lot of bureaucratic effort. This construct of "administrative capacity" is operationalised using the number of land holdings in a district as a proxy for complexity of the data gathering task. As the number of holdings increases, so does the burden on the bureaucracy to collect and collate all the data. The variable used here is the Number of Land Holdings (N_A) in the district. The following is hypothesised.

**Controlling for development, the number of land holdings in a jurisdiction is *inversely* proportional to the probability policy adoption.** This hypothesis is tested at district levels, controlling for both district and state level development indicators.

## 4 Preliminary Results and Discussion

The main results of the logistic regression are shown in Table 6. From this we see that the state's development level is extremely significant to policy adoption. Keeping other parameters at their mean, an increase in the state's development index of one standard deviation from the mean results in the log-odds of adopting the program increasing by 48%. On the other hand, the district's development level, its ruralness and number of holdings are all significant at the 10% level, but in the opposite direction. An increase in the district's development by one standard deviation from the mean, results in the log-odds of policy adoption reducing by 24%. In the case of ruralness, this effect is even more marked with a one s.d. increase in the ruralness index reducing the log-odds of policy adoption by a huge 87%. Also, the more holdings in the district, the less would it be willing to adopt the policy, with the log-odds of policy adoption dropping by 16%.

From the foregoing, while H1 is partially confirmed at the state level, however, more developed districts have a lesser propensity to adopt the policy. This looks surprising, however it is possible that administratively it is being decided to shift resources to less-developed districts and thus this result. Hypothesis 3 of administrative capacity is confirmed as we see that districts with more holdings are less liable to be selected for policy adoption. However, hypothesis 2 needs further research to understand the probability of adopting a policy meant to benefit the rural population reduces the more rural a district is.

## 5 Conclusion and Future Work

This work focuses on the challenges of creating and gathering data in the real world, or as the World Development Report 2016 says, providing the "analog foundations of our digital world" (World Bank, 2016). It is a preliminary study of the factors that potentially impact adoption of policies that aim to create data, looking at a narrow domain. This research identifies that data generation policies are hampered by not only by resource crunches, but also by the political environment, and the bureaucracy's capabilities and capacity to implement the on-ground data gathering capabilities. Further research is needed into the results thrown up by this study, especially to untangle some of the complexity that the empirical findings hint at, possibly using other methods of inquiry.

## Acknowledgements

Thanks are due to my doctoral committee members—Dr. Philip E. Auerswald, Dr. Siona Listokin and Dr. Aditya Johri for guiding me in this research. Thanks are also due to the Schar School of Policy & Government for providing the intellectual and physical space and necessary support to undertake this research.

## References

Ahuja, M. & Habibullah, W. (2005). *Computerisation of Land Records*. In Land Reforms in India. SAGE Publications India Pvt., Ltd.

Borne, K. D. (2013, November). What is Data Science and Why is it Needed? Learning From Data, Big and Small. Retrieved from http://complex.gmu.edu/www-phys/colloquium/Fall_2013/Fall%5C%202013%5C%20abstracts/kborne-SPACS-2013nov14.pdf

Dale, P. & Mclaughlin, J. D. (1999). *Land Administration*. New York: Oxford University Press.

Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., … Chuang, I. (2014, September). Privacy, Anonymity, and Big Data in the Social Sciences. *Commun. ACM*, *57*(9), 56–63. doi:10.1145/2643132

Decker, P. T. (2014). Presidential Address: False Choices, Policy Framing, and the Promise of "Big Data". *Journal of Policy Analysis and Management*, *33*(2), 252–262. doi:10.1002/pam.21755

Deininger, K. & Goyal, A. (2012). Going digital: credit effects of land registry computerization in India. *Journal of Development Economics*, *99*(2), 236–243. doi:10.1016/j.jdeveco.2012.02.007

Desouza, K. C. & Jacob, B. (2014). Big Data in the Public Sector: Lessons for Practitioners and Scholars. *Administration & Society*. doi:10.1177/0095399714555751. eprint: http://aas.sagepub.com/content/early/2014/11/06/0095399714555751.full.pdf+html

Feder, G. & Feeny, D. (1991, January). Land Tenure and Property Rights: Theory and Implications for Development Policy. *The World Bank Economic Review*, *5*(1), pp. 135–153. Retrieved from http://www.jstor.org/stable/3989973

Feder, G. & Nishio, A. (1998). The benefits of land registration and titling: economic and social perspectives. *Land Use Policy*, *15*(1), 25–43. doi:10.1016/S0264-8377(97)00039-2

Hilbert, M. (2013). Big Data for Development: From Information-to-Knowledge Societies. *Available at SSRN 2205145*. doi:10.2139/ssrn.2205145

Karch, A. (2007). Emerging issues and future directions in state policy diffusion research. *State Politics & Policy Quarterly*, *7*(1), 54–80. doi:10.1177/153244000700700104. eprint: http://spa.sagepub.com/content/7/1/54.full.pdf+html

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Eds.). (2014). *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. Cambridge University Press.

Laney, D. (2001, February). 3-D Data Management: Controlling Data Volume, Velocity and Variety. Retrieved from http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/

Lang, A. G. (1981). Computerised Land Title and Land Information. *Journal of Law and Information Science*, *1*(2), 230–255.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205. doi:10.1126/science.1248506. eprint: http://www.sciencemag.org/content/343/6176/1203.full.pdf

Maggs, P. B. (1973). Automation of the Land Title System. *American University Law Review*, *22*(2), 275–331.

Makse, T. & Volden, C. (2011). The role of policy attributes in the diffusion of innovations. *Journal of Politics*, *73*(1), 108–124.

MGI. (2001, August). *India: The growth imperative*. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/india/growth_imperative_for_india

Ministry of Finance, Government of India. (2013, September). *Report of the Committee for Evolving a Composite Development Index of States*. Retrieved from http://finmin.nic.in/reports/Report_CompDevState.pdf

National Research Council. (1980). *Need for a Multipurpose Cadastre*. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog/10989/need-for-a-multipurpose-cadastre

National Research Council. (2007). *National Land Parcel Data: A Vision for the Future* (Committee on Land Parcel Databases, Ed.). ISBN: 0-309-11031-9, 172 pages, 6 x 9, (2007). Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog/11978.html

Navratil, G. & Frank, A. U. (2004). Processes in a cadastre. *Computers, Environment and Urban Systems*, *28*(5), 471–486. Cadastral Systems {III}. doi:10.1016/j.compenvurbsys.2003.11.003

Nicholson-Crotty, S. & Carley, S. (2016). Effectiveness, implementation, and policy diffusion: or “can we make that work for us?” *State Politics & Policy Quarterly*, *16*(1), 78–97. doi:10.1177/1532440015588764. eprint: http://spa.sagepub.com/content/16/1/78.full.pdf+html

Payne, G. (2004). Land tenure and property rights: an introduction. *Habitat International*, *28*(2), 167–179. Land Tenure and Property Rights. doi:10.1016/S0197-3975(03)00066-3

Schintler, L. A. & Kulkarni, R. (2014). Big Data for Policy Analysis: The Good, The Bad, and The Ugly. *Review of Policy Research*, *31*(4), 343–348. doi:10.1111/ropr.12079

Sugiyama, N. B. (2008). Theories of Policy Diffusion: Social Sector Reform in Brazil. *Comparative Political Studies*, *41*(2), 193–216. doi:10 . 1177 / 0010414007300916. eprint: http://cps.sagepub.com/content/41/2/193.full.pdf+html

Taylor, L. & Broeders, D. (2015). In the name of Development: power, profit and the datafication of the global South. *Geoforum*, *64*, 229–237. doi:http : / / dx . doi . org / 10 . 1016 / j . geoforum.2015.07.002

Taylor, L., Cowls, J., Schroeder, R., & Meyer, E. T. (2014). Big Data and Positive Change in the Developing World. *Policy & Internet*, *6*(4), 418–444. doi:10.1002/1944-2866.POI378

Törhönen, M.-P. (2004). Sustainable land tenure and land registration in developing countries, including a historical comparison with an industrialised country. *Computers, Environment and Urban Systems*, *28*(5), 545–586. Cadastral Systems {III}. doi:10 . 1016 / j . compenvurbsys.2003.11.007

UN Global Pulse. (2012, May). Big Data for Development: Opportunities & Challenges. Retrieved from http://www.unglobalpulse.org/projects/BigDataforDevelopment

Ward, J. S. & Barker, A. (2013). Undefined By Data: A Survey of Big Data Definitions. Retrieved from http://arxiv.org/pdf/1309.5821.pdf

Williamson, I. & Ting, L. (2001). Land administration and cadastral trends – a framework for re-engineering. *Computers, Environment and Urban Systems*, *25*(4–5), 339–366. doi:10 . 1016/S0198-9715(00)00053-3

World Bank. (2016). *Digital Dividends: World Development Report, 2016*. The World Bank Group. doi:10.1596/978-1-4648-0728-2

<table>
<tr><th rowspan="2">Year</th><th colspan="2">Number of</th></tr>
<tr><th>States ($N = 29$)</th><th>Districts ($N = 613$)</th></tr>
<tr><td>2009</td><td>17</td><td>64</td></tr>
<tr><td>2010</td><td>10</td><td>68</td></tr>
<tr><td>2011</td><td>12</td><td>67</td></tr>
<tr><td>2012</td><td>9</td><td>59</td></tr>
<tr><td>2013</td><td>5</td><td>64</td></tr>
<tr><td>2014</td><td>7</td><td>78</td></tr>
<tr><td>Total</td><td>26</td><td>400</td></tr>
</table>

Table 1: NLRMP Proliferation over the years

<table>
<tr><th>Variable</th><th>Hypothesis</th><th>Type</th><th>Description</th></tr>
<tr><td rowspan="3">NLRMP</td><td>Hypothesis I. Resource Availability</td><td rowspan="3">Dependent Variable</td><td rowspan="3">Indicator of “Policy Adoption” This is a dichotomous variable that takes the value of “1” if policy is adopted and “0” for no-adoption.</td></tr>
<tr><td>Hypothesis II. Constituency Effects</td></tr>
<tr><td>Hypothesis III. Administrative Capacity</td></tr>
<tr><td rowspan="4">DevIDX_ST</td><td>Hypothesis I. Resource Availability (State Level)</td><td>Predictor</td><td rowspan="4">A State Level Development Index. This is taken from the report of the Raghuram Rajan Committee setup in 2013 to evolve a composite state development index (Ministry of Finance, Government of India, 2013).</td></tr>
<tr><td>Hypothesis I. Resource Availability (District Level)</td><td>Control</td></tr>
<tr><td>Hypothesis II. Constituency Effects</td><td>Control</td></tr>
<tr><td>Hypothesis III. Administrative Capacity</td><td>Control</td></tr>
<tr><td rowspan="3">DevIDX</td><td>Hypothesis I. Resource Availability (District Level)</td><td>Predictor</td><td rowspan="3">A District Level Development Index. This has been computed using Principal Components Analysis of district level statistics as provided by the 2011 Census and the Niti Aayog</td></tr>
<tr><td>Hypothesis II. Constituency Effects</td><td>Control</td></tr>
<tr><td>Hypothesis III. Administrative Capacity</td><td>Control</td></tr>
<tr><td rowspan="3">Ruralness_IDX</td><td>Hypothesis I. Resource Availability</td><td>Predictor</td><td rowspan="3">This is a measure of how rural and agriculturally dependent a district is. It is created using Principal Components Analysis of variables that are highly correlated (see Tables 4 and 5).</td></tr>
<tr><td>Hypothesis II. Constituency Effects</td><td>Predictor</td></tr>
<tr><td>Hypothesis III. Administrative Capacity</td><td>Predictor</td></tr>
<tr><td rowspan="3">N_A</td><td>Hypothesis I. Resource Availability</td><td>Predictor</td><td rowspan="3">This variable is the Number of Agricultural Land Holdings in the district. It is used as a proxy for the complexity of policy implementation, and thus “administrative capacity”.</td></tr>
<tr><td>Hypothesis II. Constituency Effects</td><td>Predictor</td></tr>
<tr><td>Hypothesis III. Administrative Capacity</td><td>Predictor</td></tr>
</table>

Table 2: List of Variables

Table 3: Summmary Statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| NLRMP | 613 | 0.653 | 0.477 | 0 | 1 |
| DevIDX_ST | 613 | 0.417 | 0.159 | 0.210 | 0.950 |
| DevIDX | 613 | 0.544 | 0.180 | 0.000 | 0.970 |
| Area_LQ_Rural | 613 | 0.994 | 0.051 | 0.650 | 1.230 |
| TWFRAgri_LQ_Tot | 613 | 1.068 | 0.307 | 0.090 | 2.780 |
| N_A | 613 | 224,820.000 | 188,289.900 | 1,246 | 982,314 |

Table 4: Correlation table

| | DevIDX_ST | DevIDX | Area_LQ_Rural | TWFRAgri_LQ_Tot | N_A |
|---|---|---|---|---|---|
| DevIDX_ST | 1 | 0.742 | -0.027 | 0.137 | 0.100 |
| DevIDX | 0.742 | 1 | -0.217 | -0.166 | -0.104 |
| Area_LQ_Rural | -0.027 | -0.217 | 1 | 0.668 | 0.037 |
| TWFRAgri_LQ_Tot | 0.137 | -0.166 | 0.668 | 1 | 0.052 |
| N_A | 0.100 | -0.104 | 0.037 | 0.052 | 1 |

Table 5: Correlation table with Ruralness Index

| | NLRMP | DevIDX_ST | DevIDX | Ruralness_IDX | N_A |
|---|---|---|---|---|---|
| NLRMP | 1 | 0.074 | 0.032 | -0.047 | -0.053 |
| DevIDX_ST | 0.074 | 1 | 0.742 | 0.060 | 0.100 |
| DevIDX | 0.032 | 0.742 | 1 | -0.210 | -0.104 |
| Ruralness_IDX | -0.047 | 0.060 | -0.210 | 1 | 0.049 |
| N_A | -0.053 | 0.100 | -0.104 | 0.049 | 1 |

Table 6: Comparing the Logistic Models

| | Policy Sanctioned | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| State Dev Index | $0.158^{*}$ | $0.234^{*}$ | $0.317^{**}$ | $0.392^{***}$ |
| | (0.087) | (0.128) | (0.137) | (0.143) |
| District Dev Index | | $-0.103$ | $-0.196$ | $-0.273^{*}$ |
| | | (0.126) | (0.137) | (0.144) |
| Ruralness Index | | | $-1.875^{*}$ | $-2.009^{*}$ |
| | | | (1.076) | (1.081) |
| Number of Holdings | | | | $-0.172^{*}$ |
| | | | | (0.089) |
| Constant | $0.634^{***}$ | $0.635^{***}$ | $1.563^{***}$ | $1.633^{***}$ |
| | (0.085) | (0.085) | (0.542) | (0.545) |
| N | 613 | 613 | 613 | 613 |
| Log Likelihood | $-394.236$ | $-393.899$ | $-392.327$ | $-390.481$ |
| AIC | 792.472 | 793.798 | 792.654 | 790.963 |

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$