

Digitizing Historical Plant Level Panel Data on Labour Outcomes

Preserving 40 years of labor micro data

Ananya Kotia (London School of Economics)

Utkarsh Saxena (University of Oxford)

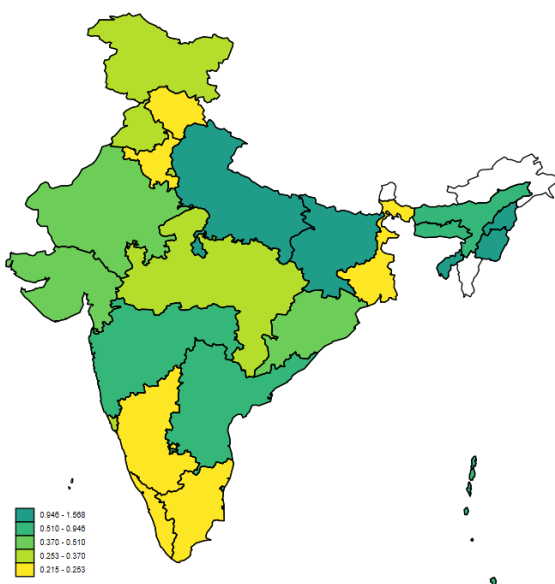


New, monthly micro data on India's industrial workers to shed light on labour turnover and absenteeism over forty years.

Follow us on Twitter! Find @glmlic

Topic at a Glance

This initiative focuses on the digitization of India's historical labor records from the Annual Survey of Industries (ASI), Part II, spanning 1960–2000. These records, previously inaccessible, offer monthly data on critical labor metrics—working days, man-days, absenteeism, turnover, and earnings. By digitizing and preserving these datasets, this project transforms them into a publicly accessible data product hosted on the Labor Bureau's website.



Caption: Variation in Worker-Firm Separations Across Months.

New Insights

The digitization of ASI Part II data provides several critical insights into India's industrial labor dynamics, offering a rich new data product that was previously unavailable. The historical records spanning four decades allow for a more nuanced understanding of employment trends, workforce behavior, and industrial seasonality. Key insights from the project include:

1. Enhanced Accessibility

This dataset, derived from the Annual Survey of Industries (ASI) - Part II, spans over 40 years and was previously inaccessible due to outdated storage formats. The digitization effort democratizes access to this vast repository, allowing researchers, policymakers, and industry stakeholders to explore workforce dynamics at an unprecedented level of detail. By making these records publicly available through government and research platforms, the dataset helps bridge long-standing data gaps in India's industrial sector, ensuring that policymakers and researchers can analyze workforce patterns over multiple decades.

2. Detailed Metrics

The monthly granularity of the dataset offers an opportunity to analyze labor trends at a finer resolution than traditional annual reports. By capturing data on working days, man-days worked, absenteeism, turnover, and wages, this dataset enables the study of short-term employment fluctuations and industrial workforce trends.

For instance, the validation of ASI Part II data for 2000-01 reveals distinct patterns in labor dynamics across India's industrial landscape. Our analysis, supported by comprehensive visualizations, identifies several key temporal and geographic patterns:

- **Seasonal Working Patterns:** Analysis of working days demonstrates clear seasonal variations, with peak activity occurring in June and July across Indian states. This pattern suggests a synchronized increase in industrial activity during these months, likely

corresponding to periods of heightened market demand and favorable working conditions.

- **Geographic Labor Demand Variations:** The choropleth mapping of man-days worked reveals significant regional disparities in labor demand stability. This geographic variation suggests different industrial compositions and market dynamics across regions.

- **Absenteeism Patterns:** The data reveals a clear seasonal pattern in workplace attendance, with peak absenteeism occurring in June across various industries. This trend strongly correlates with India's summer season, suggesting that environmental factors significantly impact workforce availability.

3. Critical for Policymaking

This dataset serves as a crucial tool for evidence-based policymaking, offering insights into workforce behaviors that were previously difficult to quantify.

This comprehensive dataset serves as both a historical record and a foundation for future planning. Its detailed metrics and long-term coverage provide invaluable insights into India's industrial development journey, while offering practical tools for current policy formation and industrial strategy development. The successful digitization and standardization of this historical data opens new avenues for research and analysis, contributing significantly to our understanding of industrial labor dynamics in developing economies.

Policy Recommendations

- **Support Data Portals:** Expand public access to this data product via the National Data Analytics Portal (NDAP) and other platforms to maximize usage.
- **Integrate into Policy Planning:** Encourage policymakers to utilize this dataset in planning industrial workforce programs, especially addressing regional disparities in labor dynamics.
- **Ensure Long-Term Preservation:** Establish protocols for sustainable storage and periodic updates to maintain the dataset's relevance and usability.

Limitations

- **Unreadable Structure:** Many of the recovered data files are in unstructured or raw formats, requiring significant preprocessing and cleaning before they can be used effectively.
- **Technical Recovery Issues:** Challenges in reading and recovering data from degraded tapes may limit the completeness of the product.

Read more → g2lm-lic.iza.org/projects/fact-and-policy/digitising-historical-plant-level-panel-data-on-labour-outcomes/

G²LM|LIC is a joint program by IZA and the UK Foreign, Commonwealth & Development Office (FCDO) – improving worldwide knowledge on labor market issues in low-income countries (LICs) and providing a solid basis for capacity building and development of future labor market policies.