Measuring Employment Using Big Data on Electronic Salary Payments

Dr. Anil Kumar Sharma*
Adviser, Department of Statistics & Information Management, Reserve Bank of India, Mumbai, India – anilksharma@rbi.org.in

Ravi Shankar
Director, Department of Statistics & Information Management, Reserve Bank of India, Mumbai, India – ravishankar@rbi.org.in

Abstract

Data on employment/unemployment are crucial inputs to policy formulation. Timely availability of reliable employment indicator adds to the efficacy of macro-economic policy. Most of the countries use regular labour force survey and or a standard household survey to measure the extant of employment in the economy e.g. Brazil, China, India. Alternatively, government and business work-sites are surveyed for detailed data on employment e.g. US. The employment measures based on these survey methods have limitation of cost, coverage and timeliness. The paper proposes an alternative approach to measure employment situation based on the salary disbursals encountered in electronic payments systems using Text Mining Technique from electronic salary payment systems in order to build the Monthly Employment Index for the formal sector in India. Further the electronic payments data on *net salary* disbursed to employees in economy could also act as a good leading indicator of *disposable income* of households that would ultimately generate further demand in the economy.

JEL Classification: E24, J33

Keywords: wages, payment methods.

Disclaimer: Views expressed are of authors and not the organisation to which they belong.

1. Introduction

Data on employment/unemployment are crucial inputs to policy formulation. In India, National Sample Survey Office (NSSO)'s quinquennial Employment-Unemployment survey is the most conventional and popular source of data on employment and unemployment. Other than this survey, presently, there are three sources of data on employment/unemployment viz. "Quarterly Quick Employment Survey", "Annual Employment Unemployment Survey" of Labour Bureau and "Annual Survey of Industries". Besides the Government sources of employment statistics, few private organisations also provide/compile employment indicators. But coverage and time-lag in release of these indicators is an issue. On the other hand, most of the organisations in organised sector in India have now started disbursing salary electronically in the account of the employee which is credited mostly through National Electronic Funds Transfer (NEFT)/ National Automated Clearing House (NACH) payment systems in the respective bank account of the employee declared by him/her to the employer. NACH system handled more than 1.4 billion retail payment transactions aggregating more than ₹ 3.8 trillion in the year 2015-16. Similarly, NEFT handled more than 1.2 billion retail payment transactions aggregating more than ₹ 83 trillion in the year 2015-16. Post-demonetisation decision in Indian economy and with emphasis on digital payments, both NACH and NEFT handled about 150 million transactions (approximately) amounting ₹ 592 million and ₹10.9 trillion, respectively, in the month of February 2017 (Table 1).

This will get further boost with recent passing of "Payment of Wages (Amendment) Bill 2017" by the parliament (Feb 2017) empowering central and state governments' to specify and allow industries to pay wages digitally, through direct bank transfers or by cheque. The bill brings an amendment to Section 6 of the Payment of Wages Act, which provide for crediting the wages in the bank account of the employee or payment through cheque along with the existing provisions of payment in current coin or currency notes.

In NEFT system, the payment message in MT Standard is sent by the Employers or debiting Bank to the Employee's Bank or the crediting account which comes for settlement to RBI on hourly basis. The message format used in NEFT is IFN298N06 which has a provision for the Sender to Receiver Information.

In this study an attempt is made to cull out salary credit data using **Text Mining Technique** from NACH and NEFT systems in order to build the Monthly Employment Index for the formal sector in India. The study is mainly focused on exploring the potential of alternative approach of measuring employment situation based on salary data encountered in various electronic payment instructions. The study is divided into five sections. Section 2 provides a backdrop of available employment information in Indian economy. The main features of major payment systems data used and its' limitations are described in the Section 3 and international practices/approaches to measure employment situations are presented in Section 4. The data mining technique used is described in Section 5. Based on the empirical data extracted using data mining, an indicator is constructed as a weighted average and the methodology used and the summary of results are presented in Section 6. Section 7 presents the conclusions and way forward for further work in this area.

2. Availability of Employment Data in India

2.1 Availability of data on Employment and Unemployment

Data on employment/unemployment are crucial inputs to monetary policy. In India, National Sample Survey Office (NSSO)'s quinquennial Employment-Unemployment survey is the most conventional and popular source of data on employment and unemployment. Other than this survey, there are three other sources of data on employment/unemployment viz. "Quarterly Quick Employment Survey", "Annual Employment Unemployment Survey" of Labour Bureau and "Annual Survey of Industries".

A. Quinquennial Survey on Employment by NSSO

The quinquennial survey on employment and unemployment is one of the important household surveys conducted regularly by the NSSO. The first such survey, in present series, was carried out by the NSSO during October 1972 - September 1973. Nine such comprehensive surveys on the employment and unemployment situation in India have so far been conducted by NSSO. The *last survey* of NSS (68th round) conducted during July 2011-June 2012 covering a large sample of 101,724 households on all-India basis. The survey provides estimates on various indicators of Employment and Unemployment in national and state level.

B. Quarterly Quick Employment Survey by Labour Bureau

The financial crisis of 2008 impacted the employment situation globally. A need was felt to have an indication about the impact of the Global Financial Crisis on employment situation in India on a quarterly basis. The Government of India, therefore, instituted the Quarterly surveys on employment changes in selected sectors by Labour Bureau. The first such survey for the quarter October-December, 2008 was undertaken during the month of January, 2009. The "Quarterly Quick Employment Survey" is conducted to assess changes in employment in selected eight important sectors such as, Textiles, Metals, Automobiles, Gems & Jewellery, Transport, IT/BPO, Leather and Handloom/Power loom. It covers around 2000-3000 enterprise units spread across various states/UTs, selected based on two stage stratified sampling technique. This survey records the number of direct and contract employees of an enterprise as well as the total wage/salary bill. Estimates of employment in each sector is derived statistically. The new series has enhanced coverage of more than 10,000 units across 18 sub-sectors. The first survey of new series was conducted in April 2016 and a report on the salient survey results is released in public domain in Jan 2017.

C. Annual Employment and Unemployment Survey by Labour Bureau

The "Annual Employment Unemployment Survey" is conducted to fill data gap in ascertaining the employment-unemployment scenario in the country. The survey was initiated in 2010 and it covers almost 6,80,000 individuals from 1,40,000 households spread across all states/UTs and all districts of India. The survey captures education level including vocational training and employment status of population aged 15 years and above. The survey results are released in five volumes namely 'Employment-Unemployment Scenario', 'Youth Employment-Unemployment Scenario', 'Education, Skill Development & Labour Force', 'Employment in Informal Sector & Conditions of Informal Employment' and 'Employment-Unemployment Scenario among different Social Groups'. The survey results are available at State/UT level, gender-wise, rural-urban level and based on NIC code in the reports. The survey is quite exhaustive in terms of coverage and availability of detailed information related to employment-unemployment and related characteristics. The latest data pertains to fifth round survey covering 2015-16 (source: labourbureaunew.gov.in/UserContent/EUS_5th_1.pdf).

D. Annual Survey of Industries

The Annual Survey of Industries (ASI) is a statutory survey conducted every year by Govt. of India, under the provisions of the Collection of Statistics Act, 1953. The survey captures data on number of workers employed directly, workers employed through contractor, supervisory and managerial staff, wages/salaries, bonus, contribution to provident fund and other funds, workmen and staff welfare expenses and various other aspects of labour statistics. The ASI extends to the entire country. It covers all factories registered under Sections 2m(i) and 2m(ii) of the Factories Act, 1948.

Data on employment based on the NSSO/Labour Bureau surveys are available in public domain. Specifically, employment data of people who received vocational training is available from Annual Employment Unemployment Survey. Further, the estimates are based on large sample surveys conducted scientifically and are derived statistically. Besides the Government sources of employment statistics, few private organisations viz., Centre for Monitoring Indian Economy (CMIE) and Naukri.com also provide/compile employment indicators.

2.2 CMIE Unemployment rate

CMIE in collaboration with Bombay Stock Exchange (BSE) provides estimate of the labour force in India and the unemployment rate in India by directly interviewing a large sample of randomly selected households, to find the employment / unemployment status of all members of over 15 years of age. The 30-day moving average is derived from a sample of about 130,000 individuals from about 39,600 households. The sample households are from the panel of households included in CMIE's Consumer Pyramids survey. CMIE's Consumer Pyramids panel of households contains over 158,624 households and these include over 522,000 members who are over 15 years old. This is one of the largest sample of individuals from whom data is gathered on the employment and unemployment status (source: http://unemploymentinindia.cmie.com).

2.3 Naukri Job Speak Index

Naukri.com is India's one of leading job site with more than 1,25,000 jobs live at any point in time and over 38 million CV's, Naukri.com. This wide spread access to hiring activity across sectors, regions and functional areas gives the company a substantial base to build its data analysis. On the basis of job listings added to the site every month, the data shows an increase or decrease in the index. To calculate the index, job listings added to the site in July 2008 have been taken as 1000. The subsequent months have been indexed with data of July 2008. The monthly report shows hiring trends across industry sectors, geography and functional areas. There might be high volatility for certain fringe cases like smaller cities, niche industries etc. owing to a small base, but more than 48000 clients using Naukri.com leads to high reliability of the data. (Source: http://www.infoedge.in/media-press-releases.asp)

3. Various Electronic Payment Systems in India used for Salary Disbursal

Most of the organisations in India have migrated to electronic fund transfer of the salary disbursal. These organisations have adopted various electronic Payment Systems implemented by Reserve Bank of India and National Payment Corporation of India (NPCI).

3.1 Electronic Clearing System (ECS)

ECS is an electronic mode of payment / receipt for transactions that are repetitive and periodic in nature. ECS is used by institutions for making bulk payment of amounts towards distribution of dividend, interest, salary, pension, etc., or for bulk collection of amounts towards telephone / electricity / water dues, cess / tax collections, loan instalment repayments, periodic investments in mutual funds, insurance premium etc. Essentially, ECS facilitates bulk transfer of monies from one bank account to many bank accounts or vice versa. ECS includes transactions processed under National Automated Clearing House (NACH) operated by NPCI. Primarily, there are two variants of ECS - ECS Credit and ECS Debit. ECS Credit enables payments of dividend, interest, salary, pension, etc.,

3.2 National Automated Clearing House (NACH)

National Payments Corporation of India (NPCI) has implemented "National Automated Clearing House (NACH)" on December 29, 2012 for Banks, Financial Institutions, Corporates and Government a web based solution to facilitate interbank, high volume, electronic transactions which are repetitive and periodic in nature. NACH System can be used for making bulk transactions towards distribution of subsidies, dividends, interest, salary, pension etc. and also for bulk transactions towards collection of payments pertaining to telephone, electricity, water, loans, investments in mutual funds, insurance premium etc.

From August 2016, NPCI has introduced a distinct NACH Credit product for the purpose of crediting employee salary accounts. This product enables destination banks and accounts to identify salary credits received and prioritise the same. With this distinct product, the data could easily be

used for compiling analysis on salary payments by the corporates, banks and other users with the help of this product identifier.

3.3 National Electronic Fund Transfer (NEFT)

National Electronic Funds Transfer (NEFT) is a nation-wide payment system facilitating one-to-one funds transfer. Under this Scheme, individuals, firms and corporates can electronically transfer funds from any bank branch to any individual, firm or corporate having an account with any other bank branch in the country participating in the Scheme. Individuals, firms or corporates maintaining accounts with a bank branch can receive funds through the NEFT system.

3.4 Limitations

For capturing intra-bank transactions we need to tap the core banking solutions of the Bank. So if we base employment index calculation based on Payment and Settlement data then Intra-bank Salary transactions will not be captured in index compilation.

As employer disburse *Net Salary* to their employees after their usual deductions, the amount of salary disbursed could not be used for estimating wage rate e.g. Salary per hour. Further, since both NEFT and NACH mostly cater to formal/organised sector, it may not capture the employment scenario in the unorganised sector which constitute major portion for an economy like India where salary is mostly disbursed through Cheque or Cash. Further it would not capture part of the employees in organised sector where salary are credited in their Staff/Officers Society accounts and not in Banks.

4. International Practices

The Bureau of Labour Statistics (BLS) handles the collection of data on employment in US (source: http://www.bls.gov/ces/). As mentioned therein, each month the Current Employment Statistics (CES) program surveys approximately 146,000 businesses and government agencies, representing approximately 623,000 individual worksites, in order to provide detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls.

Internationally, regular Labour Force Survey is a standard household-based survey of work-related statistics, e.g. Monthly employment survey of Brazil, Annual survey on labour force of China. Most of the countries use their Labour Force Surveys for measuring employment rate and survey results are available online. A wide range of labour statistics and meta-data for over 200 countries and territories are also available from the ILOSTAT database (source: http://www.ilo.org/dyn/lfsurvey/lfsurvey.home).

As seen from the above, there is no precedence of using payments information for working out employment data in any major economy.

5. Methodology for Text Mining

5.1 NEFT Data

In this study, we have used NEFT and NACH granular daily transactional data to capture the details of salary disbursals credited through these systems. For extracting salary data from NEFT system, the master data file maintained at central bank (i.e. Reserve Bank of India) is used. The following methodology is used for data-mining:

i. Date of transaction lies between 25th of the salary month to 10th of the next month.
ii. Remittance instruction includes words such as `SALARY', or `salary' or `Salary', or `SAL', or `sal', or `Sal' as characters.
iii. Sender message includes words such as `SALARY', or `salary' or `Salary', or `SAL', or `sal', or `Sal' as characters.
iv. If userid field used for salary disbursal is utilised.
(The words "pay" could also have been considered).

5.2 NACH Data

For capturing salary transactions from NACH system, the following methodology is used.

i. Transactions considered only once in the month during date range from 1st to 10th and 25th to 31st.
ii. all financial inclusion user names removed
iii. all financial inclusion user codes i.e. alpha numeric is removed
iv. Created unique user names by merging (or) deletion of names.
v. User numbers with average transaction amount <₹ 2,00,000 and > than₹ 2,500 is considered for this exercise (assuming net salary payments)
vi. Life Insurance Corporation (LIC), finance credits, and pension type transactions are not included in the data base.
vii. Ensured that the unique user name to be spread across all the months.

The above conditions/ logic for data extraction could be further fine-tuned keeping in view the situations prevailing in the economy / country e.g. data need to incorporate transactions failed or returned (return clearing).

6. Summary of Results

6.1 Empirical Data

In order to build the Employment Index for the formal sector in India, 371 common entities which distributed net salaries of more than 4.7 million in March 2016 were utilised.

Weight: Average Net Salary disbursed during the current month divided by Six Month Moving Average of Net Salary disbursed in the previous six months (i.e. Chain Base) is taken as weight.

6.2 Employment Index

Employment Index is calculated as follows:

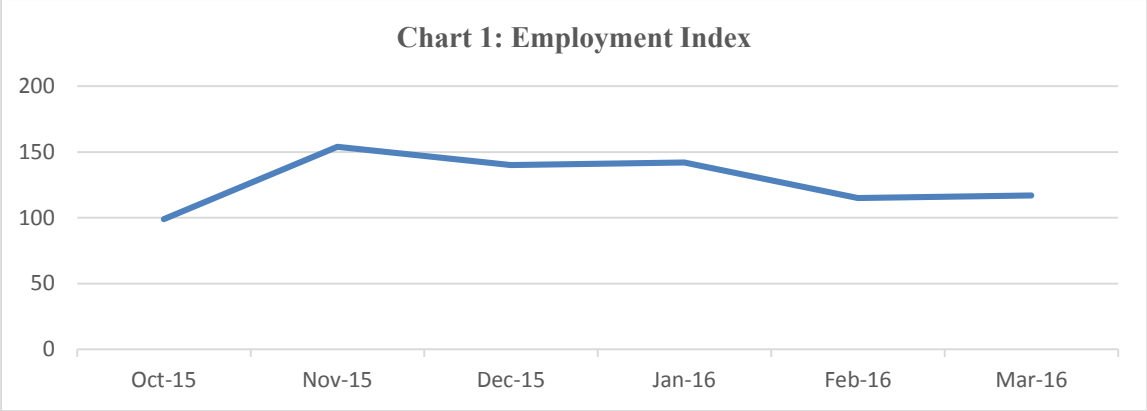
$$\text{Index value} = \frac{\text{Number of Employees in Current Month} * \text{Weight} * 100}{\text{Moving Average of Number of Employees in the Previous Six Months}}$$

Based on the above formulae employment index for the formal sector based on 371 entities covering more than 4.7 million employees is estimated as presented below.

Month	April 2015	Oct 2015	Nov 2015	Dec 2015	Jan 2016	Feb 2016	Mar 2016
Employment Index	**100**	**99**	**154**	**140**	**142**	**115**	**117**

Further refinement in the methodology may be possible going forward as NACH coverage will improve over time once NACH is implemented across the industry for salary disbursal.

As can be seen from Chart 1, employment generation in Indian economy is maximum just before and around Diwali festival time which attained peak in Nov 2015. Results compare well with Naukari.com site which has recently said that employment has risen by approx. 22 percent during March 2015 to March 2016.



Source: Based on 371 entities covering 4.7 million employees in March 2016 who were disbursed salary through NECS/NACH system of NPCI during April 2015-March 2016.

7. Conclusions

Salary disbursal data collected from payment systems covered 4.7 million employees from 371 common entities over the period April 2015-March 2016 and the data was available on monthly frequency with a minimum lag. As can be seen from the literature review carried out on the Employment Statistics compiled in India, the coverage was the maximum among all the prevailing surveys on employment conducted in India. Since there are about 42 million persons employed in the organised sector in India and electronic salary payment data has good coverage, it would be a better source to compile Employment Index for the formal sector in India in future. NPCI has introduced separate product code for capturing salary transactions from NACH system from August 2016 onwards which will further facilitate extraction of salary data from NACH system. However, it may be pertinent to note some major limitations in this measure using big data e.g. irregular data from some employer, issues in identifying employer using names, etc., Due to such issues, the index value may sometime has large variations making it impractical. These issues could possibly be overcome to some extent through mutual interactions among concerned going forward. Net salary disbursed to employees in India through the electronic payment system could also act as a good leading indicator of disposal income which is received in the hands of Indian households which ultimately generates consumption demand in the Indian economy. Based on the experience and empirical evaluations, the methodology to construct indices using alternative weighting diagrams would lead to fine-tuning of methodology going forward. The empirical exercise presents the potential scope of this approach to measure employment situation based on salary payments data despite the known limitations of such data.

References

1. CTS Circulars : http://www.npci.org.in/CTSCirculars.aspx
2. NACH Circulars : http://www.npci.org.in/NACH_Circulars.aspx
3. Reserve Bank of India Annual Report 2015-16.
4. NSSO's Employment-Unemployment Survey- various reports published by NSSO.
5. Quick Employment Survey – various reports published by Labour Bureau, Govt. of India.
6. Annual Employment Unemployment Survey– various reports published by Labour Bureau, Govt. of India.
7. Annual Survey of Industries – various reports published by CSO, MoSPI, Govt. of India.
8. CMIE-BSE Unemployment rate- http://unemploymentinindia.cmie.com/
9. Current Employment Statistics in US (source: http://www.bls.gov/ces/).
10. Labour Statistics and Metadata of over 200 countries – ILOSTAT database (source: http://www.ilo.org/dyn/lfsurvey/lfsurvey.home).
11. Naukri Job Speak Index by Infoedge source: http://www.infoedge.in/media-press-releases.asp
12. Budget speech 2017-18: http://indiabudget.nic.in/ub2017-18/bs/bs.pdf

Table 1: Electronic Payment Systems Data - Recent Trends

Volume in million, Value in Rs. billion

Data for the period	RTGS		NEFT		CTS*		IMPS*		NACH*		UPI*		USSD**		Debit and Credit Cards at POS &		PPI #		Mobile Banking	
	volume	value	volume	value	volume	value	volume	value	volume	value	volume	value	volume (in thousand)	value (in Rs. thousand)	volume	value	volume	value	volume	value
2012-13	69	6,76,841	394	29,022	275	21,780	1	4	-	-	-	-	-	-	-	-	-	-	-	-
2013-14	81	7,34,252	661	43,786	591	44,691	15	96	87	215	-	-	-	-	-	-	-	-	-	-
2014-15	93	7,54,032	928	59,804	965	66,770	78	582	340	1,221	-	-	-	-	-	-	-	-	-	-
2015-16	98	10,35,552	1,253	83,273	958	69,889	221	1,622	1,404	3,802	-	-	-	-	-	-	748	488	389	4,041
Aug-16	9	98,592	119	8,764	82	5,925	34	268	153	682	-	-	-	-	-	-	96	56	72	1,039
Sep-16	8	1,10,564	120	9,880	79	5,737	36	289	157	590	-	-	-	-	-	-	97	56	73	1,042
Oct-16	9	97,554	133	9,505	82	5,974	42	344	169	768	-	-	-	-	-	-	127	60	78	1,135
Nov-16	8	78,479	123	8,808	87	5,419	36	325	153	607	0	1	7	7,303	206	352	59	13	72	1,245
Dec-16	9	84,096	166	11,538	130	6,812	53	432	199	627	2	7	102	1,03,718	311	522	88	21	70	1,366
Jan-17	9	77,486	164	11,355	118	6,618	62	491	159	541	4	17	314	3,81,760	266	481	87	21	65	1,207
Feb-17	9	74,219	148	10,878	100	5,994	60	482	150	592	4	19	225	3,57,055	212	391	78	19	56	1,080

Notes:

1. Data is provisional.
2. *: Source is NPCI.
3. **:Figures Negligible, Source is NPCI
4. &: Card transactions of four banks.
5. #: For Nov 2016 to Feb 2017, PPI issued by 8 non-bank issuers for goods and services transactions only.
7. Mobile Banking figures are taken from 5 banks for Nov 2016 to Feb 2017. The total volume & value of electronic payment systems does not include mobile banking.

-' Data not available

RTGS – Real time gross settlement

NEFT – National electronic funds transfer

CTS – Cheque truncation system

IMPS – Immediate payment service

NACH – National automated clearing house

UPI - Unified Payments Interface

USSD - Unstructured Supplementary Service Data

POS – Point of sale

PPI – Prepaid payment instrument