Discussion

Casual Analysis of Unreliable Data for Strategising Child Development – D Narayana

This has reference to the paper titled Securing the Future: Strategizing Child Development in Karnataka in the Aftermath of COVID-19 (Vol. 2, No. 6) by Arnab Mukherji in the latest issue of your esteemed journal. I wish to point out a few infirmities in arguments and reliability of the data, as also errors in presentation.

1. Re: Karnataka's Economy around the pandemic

Figure 3 presents the data on unemployment rate from September 2019 to September 2021. Except for April-June 2020, the unemployment rates post the first Covid wave are comparable to those in 2019-20. This is surprising. Leaving the national lockdown in April-June 2020 aside, the perception is that Karnataka suffered more during the second wave. But the CPHS data does not reflect it. Google/ Apple mobility data shows that the situation was worse in Karnataka during the second wave, with foot falls in work places, retail and recreation, and transit points all falling drastically. This makes the CPHS data suspect. There is hardly a mention of the reliability of the data, or at least limitations of the data, especially when there is a debate going on about the reliability of the data (See Dreze, Pais and Rawal, Salil Sanyal, and others).

2. Re: Public Expenditure in Karnataka

The researcher writes, "there is some evidence that expenditure compression has taken place", then goes on to use the Budget Estimates, Revenue Expenditure, Capital Expenditure and Total Expenditure for 2019-20 and 2020-21, and actual receipts and expenditure during **April – December** of the two financial years. We have the Revised Estimates for the whole of 2020-21 and up to October 2021 of FY 2021-22 available on the CAG website.

I present the data here:

- Revenue Expenditure in 2020-21 is Rs 173086 crore (Rs 179776 crore BE) compared to Rs 165292 crore (Rs 181605 crore BE) in FY 2019-20. It is an increase of 4.72%.
- Capital Expenditure in FY 2020-21 is Rs 48510 crore (Rs 46512crore BE) and Rs 42962 crore (Rs 42584 crore BE) in 2019-20. It is an increase of 12.91%.
- Total Expenditure too is Rs 218927 crore (Rs 226288 crore BE) in 2020-21 compared to Rs 202821 crore (Rs 224190 crore BE) in FY2019-20. It is higher by 7.94% in 2020-21.

So, where is the expenditure compression in 2020-21?

Also, the remark "a large amount of this was committed expenditures such as interest payments, pensions, wages, and salaries, which could not be compressed "implies that essential expenditure on health and social support has been compressed. Interest payments and pension for FY 2020-21 are 87.56% and 85.26% of the BE respectively, whereas the total revenue expenditure is 103.21% of BE. Hence, that remark too is not supported by data.

3. Re: Household Narratives

The researcher says, "data from Karnataka from wave 16 – 21 (i.e., from January 2019 to December 2020) of the CPHS" has been used to "understand how incomes and expenditure have changed at the household level". And "Table 3 presents quarterly estimates for income and expenditure for the last two quarters of the financial years 2017-18, 2018-19 and 2019-20 and for the first month of 2020-21". That should be seven entries. In Table 3, however, there are only six entries, with the first month of 2020-21 missing. Yet the researcher draws the inference that "Table 3 shows that household budgets have been affected by COVID-19 – incomes declined by 14% in the January-March 2020 cohort, and further reduced by 38% in April 2020". The numbers in the Table show that incomes increased by 26% in the January-March 2020 quarter compared to the previous quarter, and by 48% compared to the previous year. So, where is the reduction? Further, how do you assess the impact of COVID-19 without providing data for the period?

4. Re: Defining Priority Areas

The analysis of this section is based on Table 8 (not Table 1 as shown on page 50). The researcher draws inferences based on change between 2015-16 and 2019-20 in the selected indicators such as wasting, stunting etc., of children. If we go by the 'falling behind' computed by the researcher, then Bagalkot, Koppal, and Yadgir, for example are the best performers and Chikkamagaluru, Dakshina Kannada, and Kodagu are the worst performers.

However, consider the "% Stunted" figures in 2019-20: the levels are between 25.1% and 30.4% in the latter group of districts, compared to over 48% in the former group. Similar is the picture with regard to underweight, where the latter group reports levels below 27% while the former group has levels above 45%. So, if we go by the levels achieved, then the latter group of districts are the best performers. This highlights the importance of considering levels as well as change.

There is a related issue of data reliability here too. Take an indicator such as "% women 15-49 years who report BMI below normal". Chikkamagaluru, Dakshina Kannada, and Kodagu report levels between 12% and 14.2%, compared to the levels between 17% and 26% in Bagalkot, Koppal, and Yadgir. The gains of the former group are also phenomenal. But severe child wasting in the former ranges between 9.3% to 15.7%, compared to 3.8% to 7.2% in the latter. It is hard to believe that in districts where the women are so well fed the children are allowed to go 'waste'. It looks like the small sample size and estimation at the district level of NFHS is doing something well beyond our understanding. Before using the data, do we not have a responsibility to say something about their reliability, or at least to flag these issues of concern?

5. Errors in Presentation

Refer to Table 3. Source of the table is not shown. Notes to the Table mentions COV (coefficient of variation), but the Table has only Gini and not COV. The same problem gets repeated in Table 4. The Table shows data bifurcated by educational level. I am not sure column numbers and headings tally as regards income, expenditure, and residuals. Table number on page 46 and page 50 are wrong; they should have been 7 and 8 instead of 1 and 2. There are many more such examples one could find.

Response by Arnab Mukherji

This is with reference to the comments from Dr. D. Narayana on the paper titled "Securing the Future: Strategizing Child Development in Karnataka in the Aftermath of COVID-19".

Thank you for a careful read and for the comments on data and other related issues. This helps emphasize some of the points highlighted and I'd like to respond to some of them specifically, and others more generally.

1. The paper was written at a point in time when COVID-19 Wave 1 had peaked, but Wave 2 wasn't on the horizon, i.e. around November 2020. At that time, neither was the microdata from CMIE available to capture the April– June 2020 quarter and nor was the fiscal data ready beyond the Budget Estimates (B.E.). Further, the contemporaneous public expenditure data, also taken also from the CAG website, shows that by December in FY2019-20 71% of total revenues had been received, whereas it was merely 59% in FY2020-21. The analysis presented is thus based on what would be seen as the impacts associated with very early exposure. Writing the paper 6 months down the road would of course imply availability of newer data, data patterns, and different inferences.
2. The CMIE data for Wave 21, labelled as March 2020 in Table 3, is for the period January – March 2020. The CMIE data essentially captures the shock that starts with the COVID-19 crisis building up and ending in the complete lockdown of the economy by the 3rd week of March 2020. This shock extended into the subsequent financial year, but this 2–3-week exposure is all that we capture in the paper. The key insight of the paper is that this initial exposure is sufficient to generate a very significant increase in inequality in some socio-economic groups – for example, if we look at households where all members are illiterate (see Table 4), we see that inequality rises to 0.414 from 0.364 from the preceding quarter and from 0.368 from a year ago – these are very large changes. In this sense, the household shock we capture is simply the tip of the exposure, and this alone has very large economic impacts on specific sub-groups as listed in Table 4.
3. NFHS 5 data is meant to be representative at the district level and as such, there should be no difficulty estimating univariate statistics such as mean, standard deviation, or the Gini Coefficient. This is widely used as such in a similar fashion, for example, see https://geographicinsights.iq.harvard.edu/nfhs-tracker from which I quote "*... the NFHS (NFHS-4) has been the only data representative of the district level for India*"; this is also true for NFHS 5. Thus, the deficits observed are real and quite worrisome, as the comments from Dr. Narayana implicitly capture. Further, surely the fact that the rank ordering of districts on levels is different from the rank ordering of districts while measuring changes isn't surprising. Districts with the largest absolute declines need attention and investigation into processes and mechanisms that led to such declines in children's nutritional status. That fact that districts that have historically been best performers have also lost much of their gains based on datasets meant to be representative of the population should be troubling.
4. Thank you for pointing out the errors in the labelling of the table and errors in internal referencing within the paper. All inconvenience to the reader in this regard is regretted.

The primary purpose of this piece is to present a framework for prioritizing districts for intervention in the context of children's nutritional status. As our exposure and understanding of COVID19 and its economic shock changes, both the underlying data and inference will change. Hopefully, this prioritization schema to identify districts for intervention will remain useful.