Inequality and water pollution in India

Sulhi Ridzuan

Independent Researcher, Puchong, Malaysia. E-mail: sulridzuan@gmail.com

ABSTRACT

India is notorious for high inequality and high water pollution. There is a growing body of literature that says inequality is harmful to the environment, but it does not receive strong empirical support. We discuss some econometric problems that may have caused mixed findings in the empirical literature and use appropriate tools to overcome the problems. Our empirical results using Indian time-series data show (i) that inequality leads to an increase in water pollution, (ii) that the magnitude of inequality is nearly as large as that of corruption, suggesting that reducing inequality is almost as important as curbing corruption in addressing water pollution challenges in India, and (iii) that increases in water pollution, in turn, widen inequality in India. Our results are robust to various sensitivity checks. We also find no evidence of the environmental Kuznets curve hypothesis for water pollution in India.

Key words: Environmental Kuznets curve, Inequality, Time series, Water pollution

HIGHLIGHTS

- We examine the relationship between inequality and water pollution in India and take into account econometric issues in the literature.
- We find that inequality leads to an increase in water pollution and the magnitude of inequality is nearly as large as that of corruption.
- Increases in water pollution, in turn, widen inequality in India.

1. INTRODUCTION

In India, rivers are much more than bodies of water. Indian rivers are believed among the majority Hindus to be sacred, can wash away sins, and bring people closer to god. Despite perceived to be pure, the rivers are not free from pollution – they serve as a dumping ground for sewage, solid, and industrial wastes (Bari, 2018). Every day, more than 10.5 million gallons of wastewater flow into rivers and other watercourses in India (Hirani & Dimble, 2019). Water pollution in India has been posing a serious threat to the health of its economy and society. Recent research suggests that in a developing country like India, pollution upstream could cut economic growth in downstream regions by nearly a half percentage point (Damania et al., 2019). Every year, more than 400,000 Indians die from diarrheal illness due to inadequate sanitation and hygiene (DeFrancis, 2011). It is estimated that the health costs of water pollution in India amounted to about $6.7–8.7 billion per year (Mani et al., 2012). India has spent billions of dollars on clean-up efforts, yet serious water pollution still persists (The Economist, 2019a, para. 6). Given the significant impact of water pollution on the economy and society, gaining a better understanding of the underlying causes of water pollution is critical.

Water pollution in India has been attributed to an array of factors. Socioeconomic factors that are usually hypothesized to impact upon water pollution in India include income, population, urbanization, illiteracy, lack of democracy, and corruption (Karn & Harada, 2001; Goldar & Banerjee, 2004; Barua & Hubacek, 2009; Greenstone & Hanna, 2014; Sigman, 2014; Damania et al., 2019). This study explores one factor that has not been thoroughly studied for the case of India, namely the impact of inequality[1].

Our literature review presented in the next section suggests that there may be a link between inequality and water pollution in India. Little attention, however, has been paid to the question of whether these two phenomena are related. There are a few empirical studies that, though they do not explicitly provide evidence for the case of India, could offer valuable insights, but the results of these studies are mixed – with positive, negative, and insignificant relationships between inequality and water quality (see, e.g., Scruggs, 1998; Torras & Boyce, 1998; Grafton & Knowles, 2004; Clement & Meunie, 2010; Gassebner et al., 2011; Jun et al., 2011; Kasuga & Takaya, 2017).

A possible explanation for the inconsistency in results is that the empirical literature suffers from many limitations. First, some studies ignore the time-series properties of the underlying data (see, e.g., Clement & Meunie, 2010; Jun et al., 2011). Our examination of the time-series properties of inequality and water pollution (shown later in this paper) indicates that both variables are nonstationary. Regressions that involved nonstationary variables must be viewed with extreme caution as they are likely spurious (Granger & Newbold, 1974). Even if nonstationary variables are cointegrated (move together), inferences from regressions can be misleading (Stock & Watson, 2015, p. 706).

Second, some studies such as Scruggs (1998), Torras & Boyce (1998), and Grafton & Knowles (2004) use cross-sectional data. Cross-sectional regressions are well known to suffer from omitted variable bias because it is difficult to control for everything, especially for variables that are hard to measure, such as culture. Although this problem can be addressed by using panel data as is used by some studies, many panel methods assume parameter homogeneity (the effect of inequality on water pollution is assumed to be the same for all countries)[2]. If this assumption does not hold, coefficient estimates will be seriously biased (Durlauf et al., 2009, p. 617).

Third, most studies that use cross-country data ignore the potential cross-section dependence in the data, which can lead to biased statistical inference (Hoechle, 2007). Water pollution is spatially correlated because some rivers are shared by a group of countries; if one country pollutes a river, some other countries may be affected (Thompson, 2016). Fourth, many inequality data are not fully comparable across countries (Atkinson & Brandolini, 2001; Knowles, 2005). This raises question marks over the validity of findings from cross-country regressions.

Fifth, inequality data are well known to be subject to measurement error, which can seriously bias coefficient estimates (Forbes, 2000). This issue can be addressed by employing an instrumental variable estimator or a widely used generalized method of moments (GMM) estimator, such as is used in Clement & Meunie (2010). However, valid instruments are difficult to find (Wooldridge, 2013, p. 543). And GMM estimator produces inconsistent estimates if the assumption of parameter homogeneity does not hold (Durlauf et al., 2009, p. 634).

Sixth, inequality data do not evolve much in a short time period (Deaton, 2013). A sample with a small period of inequality, therefore, has limited information (due to little year-to-year variations), making it difficult to precisely estimate the effect of inequality on water pollution (Torras & Boyce, 1998; Wooldridge, 2013, p. 54). With little information, it should not be surprising that studies with a small sample (e.g., Kasuga & Takaya, 2017) find no significant impact of inequality on water pollution.

_____

[1] Much empirical work uses cross-country data, which introduces some econometric problems we present shortly.

[2] According to Cushing et al. (2015), the relationship between inequality and the environment varies across countries depending on a country's average income, development, and democracy levels.

Seventh, some previous studies do not deal with potential reverse causality or simultaneity (Scruggs, 1998; Torras & Boyce, 1998; Grafton & Knowles, 2004; Kasuga & Takaya, 2017). Not only inequality could have an effect on water pollution but also water pollution could have an effect on inequality (through health). For example, water pollution impairs poor people's health as they are more exposed to such pollution than the rich are (Ravi Rajan, 2014). Poor health subsequently increases absenteeism and reduces productivity at work, reducing the earning capacity of the poor (Cole & Neumayer, 2006). Sickness also lowers the attendance of poor children in school, affecting their educational attainment and thereby future earnings (Cole & Neumayer, 2006). Failure to account for this reverse causation (i.e., water pollution affects inequality) will result in biased estimates. Possible solutions – for example, IV and GMM estimators – have limitations we mentioned above.

This study uses the Johansen (1995) approach to overcome the above limitations. It views all variables to be endogenous, and thus allows for simultaneous relation or reverse causation between inequality and water pollution. Biases due to omission of variables are also likely to be minimal because cointegration estimates are super consistent[3] and are robust to omitted variables (Bonham & Cohen, 2001). Moreover, a cointegration relation is invariant to increases in the information set (Juselius, 2006, p. 11). It is different from standard regression analysis, where one additional variable can considerably change parameter estimates (Juselius, 2006, p. 11)[4]. As shown by Hassler & Kuzin (2009), the Johansen approach is also robust towards measurement errors. This study focuses on a single country – India – and thus is not subject to problems associated with cross-country regressions we mentioned above – parameter homogeneity, cross-sectional dependence, and data comparability. We use data for a long time span (1978–2008)[5] and focus on a long-run relationship. Hence, little year-to-year variations in inequality data should not be of concern in this study. One might argue that the sample size this study employs is small ($T = 31$). Although the sample size is important for statistical analysis, the sample span is also important to uncover a true relationship between particular variables (Lahiri & Mamingi, 1995). A sample of 30 observations over a span of 30 years produces better information than a sample of, say, 300 observations over a span of 300 days; or a sample of 300 observations over a span of 300 months (25 years)[6]. However, this study will take appropriate measures (such as by applying a small sample correction method) whenever possible to guard against bias from a small sample.

The objective of this study is to examine the relationship between inequality and water pollution in India. We have several main results. First, we find that inequality increases water pollution in India. Second, the effect of inequality on water pollution is about as large as the effect of corruption. Third, we also observe that increases in water pollution, in turn, exacerbate inequality in India. It is worth mentioning that no evidence of the environmental Kuznets curve (EKC) hypothesis is found for water pollution in India.

The rest of the article is organized as follows. Section 2 presents a literature review. Sections 3 and 4 describe data and methodology, respectively. Section 5 presents the results. Section 6 concludes, and Section 7 provides policy recommendations.

---

[3] Super consistent estimates converge to the true parameter values at a faster rate than consistent estimates do, as sample size increases.

[4] Cointegration tests can also serve as a misspecification test (Stern, 2011). A lack of cointegration indicates that additional nonstationary variables may be needed for the model (Stern, 2011). The reason is that if relevant nonstationary variables are omitted, they will be part of the error term, making the error term nonstationary (Everaert, 2011).

[5] The sample period is dictated by the availability of data.

[6] 'Whether the sample is 'small' or 'big' is not exclusively a function of the number of observations available in the sample, but also of the amount of information in the data. When the data are very informative about a hypothetical long-run relation, … we might have good test properties even if the sample period is relatively short' (Juselius, 2006, p. 141).

2. LITERATURE REVIEW

There is a growing body of theoretical literature that says inequality is harmful to the environment (Cushing et al., 2015; Islam, 2015). Several possible causal pathways have been offered. Inequality is bad for the environment because the rich consume more (e.g., water for a private swimming pool, gardening, and washing cars) and thereby generate more waste and pollution (Islam, 2015; Dorling, 2017). This hypothesis could be true in India. For example, from a survey of 1,495 households in Bengaluru by the Ashoka Trust for Research in Ecology and the Environment, it is found that rich households (top ten percentile) use four times more water than average households do; rich households consume 340 L per person per day while average households consume 85 L per person per day (The Hindu, 2017, para. 1). According to the survey, gardening and washing cars are among the reasons for the high consumption (The Hindu, 2017, para. 3). The inequality-environment theoretical literature also argues that the rich also set the norm for what defines a lifestyle of high status, inducing the rest to work more and consume more (Berthe & Elie, 2015; Cushing et al., 2015; Islam, 2015). It is well known that, in India, owning a car is considered a status symbol (Mohan, 2020). Imitating the lifestyle of the rich could result in more cars being bought unnecessarily and more water being used, for both automotive manufacturing and household consumption (i.e., car wash).

The theoretical literature also suggests that inequality erodes trust, cohesion, and cooperation, which undermine collective action to protect environmental resources (Wisman, 2011). This phenomenon can also be seen in India. For example, in Calcutta, interaction and cooperation among environmental NGOs are rare (Dembowski, 2001, p. 81). It is perceived that many of them view one another with a suspicion that others have a hidden agenda and are pursuing personal interests (e.g., fame and foreign funds) (Dembowski, 2001, p. 2). Moreover, those environmental NGOs find it easier to mobilize people from within their own social class than other classes (Dembowski, 2001, p. 82). In 2019, water in Calcutta ranked the second most unsafe after Delhi (DNA India, 2019, para. 1).

It is believed that, in the inequality-environment theoretical literature, inequality causes people to think about unemployment, status insecurity, and making the economy work again and shifts people's attention away from environmental issues (Wilkinson & Pickett, 2010, p. 263). This may be happening in India. In Calcutta, Dembowski (2001, p. 82) reported that the environment is not considered a serious issue for communities struggling for daily survival, although such people are among the most exposed and vulnerable to environmental hazards. In another city Kanpur, Do et al. (2018) find that public demand for water quality was low in the past; people were more concerned about the impact water pollution policy would have on the economy as they depended on the highly river-polluting tanning industry for jobs. In 2017, India's leather industry employs an estimated 3 million workers, which mostly come from poor and marginalized groups (Chitnis, 2017).

3. DATA

To investigate the impact of inequality on water pollution in India, this study uses aggregate annual time-series data. This paper focuses on water quality in rivers as they are the most important source of water in India (Agrawal, 1994). We employ biological oxygen demand (BOD) as our measure of water pollution. BOD measures the amount of oxygen consumed by organisms to remove organic matter. It has anthropogenic elements, which are important to capture the effect of socioeconomic variable such as inequality. Also, BOD is a common measure of water quality in the literature of income and the environment (e.g., Greenstone & Hanna, 2014; Sigman, 2014); so, using this measure allows for comparability. Data on BOD (mg/L)[7] are taken from Global

---

[7] Unpolluted rivers have a BOD below 2 mg/L; moderately polluted rivers have a BOD between 2 and 8 mg/L; severely polluted rivers have a BOD over 8 mg/L (Desbureaux et al., 2019).

Table 1. | Summary statistics.

	Mean	Median	Maximum	Minimum	Std. Dev.
BOD	6.17	5.21	14.80	3.81	2.41
Theil	0.08	0.08	0.11	0.07	0.01
Gini	42.64	41.50	47.20	39.80	2.41
Real GDP per capita	638.80	563.75	1,156.93	373.83	226.07

Environmental Monitory System (GEMS) (available at https://gemstat.org/). This study uses two different measures of inequality – Gini and Theil index. Data on Gini are drawn from the Standardized World Income Inequality Database (SWIID, version 7.1). As shown later in the next section, Gini from the SWIID is not robustly $I(1)$, so we will use the Theil index from the UTIP-UNIDO (available at https://utip.lbj.utexas.edu/data.html) in our main analysis and Gini for robustness checking. The main advantage of the Theil index is that it is based on datasets that are collected consistently over long intervals (and thus comparable across time) (Galbraith, 2009). This advantage is significant in the present study, given that our analysis is based primarily on long time-series data. Another advantage of this Theil index is that it is not imputed like the Gini of SWIID, and thus, is more reliable. The Theil index has been commonly used in many studies to measure country-level income inequality (see, e.g., Navarro *et al.*, 2006; Krieger & Meierrieks, 2019). Real income per capita comes from the World Development Indicators (WDI). The sample period is 1976–2008. The end year is the latest period for which all the data are available at the time of writing. As is standard, income is log-transformed. For water pollution and inequality, because one of our main purposes is to test for cointegration – to see whether both time series move together – we prefer both series to appear in their original form[8]. Table 1 provides summary statistics.

4. METHODOLOGY

The Johansen (1995) approach[9] is then employed to test for cointegration between the series. Main advantages of this approach over other common cointegration methods are that (i) it views all variables to be endogenous, (ii) it allows us to perform hypothesis testing on the cointegrating vectors, and (iii) it is robust towards measurement errors (Hassler & Kuzin, 2009). To use the Johansen approach, we estimate the following vector error correction model (VECM)

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \tag{1}$$

where y_t is an $n \times 1$ vector of the first-order integrated $[I(1)]$ variables that can be endogenous. The matrix Π contains two $N \times N$ matrices, $\Pi = \alpha\beta'$. The matrix α contains the speed of adjustment, and the matrix β' contains the cointegrating coefficients. Γ_i is an $n \times 1$ matrix of short-run parameters, and ε_t is an $n \times 1$ vector of white-noise disturbance terms. Δ is a difference operator. The above equation is estimated using maximum likelihood (ML).

To test for cointegration, we will look at the rank (r) of Π through its eigenvalues. If all the eigenvalues are not significantly different from zero (i.e., the rank of Π is zero), then there is no cointegration (variables under

---

[8] With data transformation, a relationship between variables may be found even though they are not correlated (Gujarati & Porter, 2009, p. 395).

[9] Interested readers are referred to Schmith *et al.* (2012) for a detailed explanation and an example of application of this method.

consideration do not move together in the long run). If the rank equals one, they are cointegrated (move together). There are two test statistics, which can be used to test the number for cointegration rank under the Johansen approach – the trace and the maximum-eigenvalue statistics – which are computed as

$$\lambda_{Trace}(r) = -T \sum_{i=r+1}^{n} \ln{(1 - \hat{\lambda}_i)}$$

and

$$\lambda_{max}(r, \ r + 1) = -T \ln{(1 - \hat{\lambda}_{r+1})}$$

where r is the number of cointegrating vectors, T is the number of observations, and $\hat{\lambda}_i$ is the estimated value for the eigenvalues from the Π matrix. We will first test a null hypothesis of no cointegrating vectors ($r = 0$). If the null is rejected, the null that there is at least one cointegrating vector ($r \leq 1$) will be tested. If this null is not rejected, we will conclude that there is a long-run relationship between a set of variables. Osterwald-Lenum (1992) provides critical values for the trace and the maximum-eigenvalue statistics.

5. RESULTS

To use the Johansen approach, all variables must be integrated of order 1, denoted as $I(1)$. A time series is said to be $I(1)$ if it needs differencing once to induce stationarity. A time series is stationary or $I(0)$ if its mean, variance, and covariance are constant over time. To test the order of integration of time series under consideration, we use the standard augmented Dickey–Fuller (ADF) (1979) and Phillips and Perron (PP) (1988) unit root tests. Table 2 reports the results of the tests. In levels, the tests do not reject the null hypothesis of a unit root. For the first differences, the null hypothesis is rejected for all series except for Gini. It can be concluded that BOD, Theil, and income are integrated of order 1, $I(1)$. Because we cannot find strong evidence that Gini is $I(1)$, it will only be used for robustness checks.

To test for cointegration between inequality and water pollution, Equation (1) is employed on our three k-variable models. Model 1 includes BOD and Theil. Model 2 consists of BOD, Theil, and income. Model 3 includes BOD and Gini. Table 3 presents the cointegration test results. For model 1, both the trace and maximum-eigenvalue tests indicate that BOD and inequality are cointegrated. For model 2, following Heerink *et al.* (2001), we add income in our set of variables. It is common in the literature to add squared of income to account

Table 2. | Unit root tests.

	Level		First difference	
	ADF	PP	ADF	PP
BOD	−2.854800	−2.802497	−5.996425**	−7.272027**
Inequality (Theil)	−2.091449	−2.140890	−4.207627**	−5.116411**
Inequality (Gini)	−1.669229	−1.082935	−3.751055**	−3.140159
Income	−1.647986	−1.562229	−6.270841**	−7.832962**

Note: All estimates include constant and trend. The Schwarz criterion is used to determine the optimal number of lags in the ADF test. The Newey-West method and the Bartlett kernel are used for the bandwidth in the PP test.
**Significance at the 5% level.

Table 3. | Johansen's cointegration test.

	Model 1		Model 2		Model 3	
	Trace statistic	Max-eigen statistic	Trace statistic	Max-eigen statistic	Trace statistic	Max-eigen statistic
$r = 0$	19.61**	16.60**	30.47**	17.55	17.07**	17.03**
$r \leq 1$	3.01	3.01	12.92	7.01	0.048	0.048

Note: The optimal number of lags is chosen based on the Schwarz information criterion. Model 1: $y_t = (\mathrm{BOD}_t, \mathrm{Theil}_t)'$, model 2: $y_t = (\mathrm{BOD}_t, \mathrm{Theil}_t, \mathrm{Income}_t)'$, and model 3: $y_t = (\mathrm{BOD}_t, \mathrm{Gini}_t)'$.
**Significance at the 5% level.

for a nonlinear relationship between income and pollutants, known as the EKC (Grossman & Krueger, 1995). However, it is inappropriate to include the square of income in our model because, like many popular cointegration methods, the Johansen approach needs different asymptotic theory (see Müller-Fürstenberger & Wagner, 2007). Nevertheless, the EKC can still be assessed by comparing the long- and short-run coefficients on income – if the former is negative or smaller than the latter, then the EKC holds (Narayan & Narayan, 2010). The trace statistic suggests that water pollution, inequality, and income are cointegrated. The maximum-eigen-value test, however, does not confirm this result. A possible explanation for this conflicting result is that income may not belong to the cointegration relation (see Barua & Hubacek, 2009; Greenstone & Hanna, 2011). To check if this is the case, we perform long-run exclusion tests to see whether income can be excluded from our model. Table 4 also reports the results of exclusion tests. It shows that income can be safely excluded from the long-run relation (insignificant *p*-value). Given that adding the income variable does not improve our specification, our preferred model is a bivariate one[10]. Model 3 replaces Theil with Gini as a measure of inequality. As can be seen, both statistics support the presence of cointegration.

Although we find robust evidence of cointegration between water pollution and inequality, it could be argued that the results are due to the sample size we use. In small samples, the Johansen test tends to reject the no cointegration null (Cheung & Lai, 1993). To see whether our results are biased by a small sample, we adjust our test statistics for model 1 (our preferred model), by multiplying the statistics with a small sample correction factor of Reinsel & Ahn (1992), $(T - k \times n)/T$, where T is the number of observation, k is the number of lags, and n is the number of variables. This study also applies the Bartlett correction proposed by Johansen (2002) to the trace test statistic. Both corrections will lower the test statistics, making it difficult to reject the null hypothesis. Table 5 presents the cointegration test results with small sample correction. All the corrected statistics reject the null of no cointegration at the 5% level. This suggests that our finding of cointegration is not driven by our small sample.

Table 6 presents estimates of the long-run relationship from the above three models. All models show a significant positive effect of inequality on water pollution. The estimated parameters imply that a 1% increase in inequality leads to, on average, a 1.931–2.958% increase in water pollution in the long run. The income variable in model 2 is not statistically significant. To see the existence of EKC, in unreported analysis, we compare this long-run coefficient with its short-run coefficient, as suggested by Narayan & Narayan (2010). We observe that the long run (0.655) is smaller than the short run (8.173), so no evidence of EKC. Like the long-run coefficient, the short-run coefficient is not significant, either (available upon request). This study agrees with the findings of

---

[10] Although a bivariate model is unusual in standard regression analyses, it is not the case in cointegration studies (see, e.g., Schmith *et al.* 2012; Herzer 2020). Cointegration property is invariant to increases in the information set (Juselius, 2006, p. 349). If cointegration is found within a set of variables, the same cointegration relation will still be found if additional variables are added (Juselius, 2006, p. 349).

Table 4. | Long-run exclusion tests.

Variable	BOD	Inequality	Income
$\chi^2(1)$	8.00	6.56	0.09
(*p*-values)	(0.0046)	(0.0104)	(0.7656)

Table 5. | Johansen's cointegration test with small sample correction

Correction method	Reinsel and Ahn Trace statistic	Reinsel and Ahn Max-eigen statistic	Johansen Trace statistic
$r = 0$	18.26**	15.46**	16.277**
$r \leq 1$	2.80	2.80	1.863

Note: The optimal number of lags is chosen based on the Schwarz information criterion.
**Significance at the 5% level.

Table 6. | Estimates of the effect of inequality on water pollution.

	Model 1	Model 2	Model 3
Inequality	2.146** (5.43)	1.931** (4.24)	2.958** (2.59)
Income		0.107 (0.42)	

**Statistically significant at 5%. *t*-statistics are in parentheses. All coefficient estimates are converted to elasticities for comparability.

other studies (Barua & Hubacek, 2009; Greenstone & Hanna, 2011), in which no EKC relationship for water pollution in India.

To test the robustness of this result, we re-estimate the long-run relationship using three alternative estimation methods: the dynamic ordinary least squares (OLS) of Stock & Watson (1993), the fully modified OLS estimator of Phillips & Hansen (1990), and the canonical cointegrating regression of Park (1992). Table 7 presents estimation results using these estimators. The ML estimation of Johansen (model 1; Table 6) is reported for comparison. We continue to find a positive and statistically significant effect of inequality on water pollution, and the estimates from the three methods are generally rather close to that of ML. We conclude that our results are robust to different estimators.

Next, we examine the sensitivity of our results to the inclusion of other explanatory variables. To this end, we use the dynamic OLS estimator as it has important econometric advantages over other estimators. First, it

Table 7. | Estimates of the long-run effect of inequality on water pollution.

ML	DOLS	FMOLS	CCR
2.146** (5.43)	1.922** (4.26)	1.745** (4.23)	1.747** (4.22)

ML, maximum likelihood; DOLS, dynamic OLS; FMOLS, fully modified OLS; CCR, canonical cointegrating regression. The DOLS regression is estimated with one lead and one lag.
**Statistically significant at 5%. *t*-statistics are in parentheses. All coefficient estimates are converted to elasticities for comparability.

Table 8. | DOLS estimates with corruption, democracy, and literacy.

	(1)	(2)	(3)	(4)
Inequality	1.498** (3.51)	1.724** (3.06)	1.796** (3.33)	3.042** (4.69)
Corruption	1.591** (2.49)			1.017 (1.47)
Democracy		8.452 (0.66)		− 1.027 (− 0.46)
Literacy			0.152 (0.45)	− 1.494 (− 2.03)
Sample period	1986–2007	1980–2007	1980–2007	1986–2007

t-statistics in parenthesis. **Statistically significant at 5%. Given small sample, only one lead and one lag of the regressors are used in the estimation. All coefficient estimates are converted to elasticities for comparability.

performs well in small samples as ours. Moreover, one variable that we are going to use has short time-series observations, reducing our already small sample. Second, it allows $I(0)$ independent variables (such as corruption) in the regression. Third, it is robust to endogenous explanatory variables – an issue that may arise from measurement error of inequality and reverse causation between inequality and water pollution.

Table 8 reports our results. In column 1, corruption is added as an explanatory variable. The inclusion is motivated by the fact that corruption is correlated with inequality (Gupta *et al.*, 2002), and as a result, both variables could proxy one another. There is a possibility that what we are estimating so far is the impact of corruption rather than that of inequality. The corruption data are obtained from the International Country Risk Guide. The data have been rescaled so that greater values correspond to more corruption. Column 2 adds democracy as it also tends to be correlated with inequality (Eriksson & Persson, 2003) and has been found important in explaining variation in water pollution (Lin & Liscow, 2013). The Polity2 Index, which is from the Polity IV dataset, is used to measure democracy. In column 3, we add literacy, which relates to inequality (Torras & Boyce, 1998) and has been found significant in Barua & Hubacek (2009). Literacy data for India are available from WDI but have many data gaps. For that reason, we use secondary school enrollment rate (which has far fewer missing data) from the same source as a proxy for literacy[11]. In column 4, all additional explanatory variables are entered in the regression simultaneously. Overall, the estimated effect of inequality is not very sensitive to which specific control variables are entered in the regression. In all cases, the coefficient of inequality remains statistically significant. The democracy and literacy variables are not significant. Corruption is statistically significant at 5% with expected signs in column 1[12]. In column 4, corruption is not significant, which is understandable, since we lost more degrees of freedom with more variables included. Since only one variable is significant in column 4, results in column 1 are preferred. Interestingly, we observe that the magnitude of inequality (1.498%) is nearly as large as that of corruption (1.591%), suggesting the importance of inequality in explaining water pollution[13]. It is important to bear in mind that this finding is based on small sample size. However, the estimated effect of corruption is within the range of a previous study – −0.681 to −1.682% (see Table 2 in Sigman (2014))[14] – and that gives us some confidence about the reliability of our finding. Because inequality

---

[11] Data for missing years are interpolated.

[12] Cointegration test with corruption included is not undertaken given the fact that corruption is not persistent or $I(0)$ (Seldadyo & De Haan, 2011). Furthermore, from our unit root tests, we also do not find strong evidence that corruption is $I(1)$.

[13] Note that although inequality has a relatively larger *t*-statistic than corruption does, such statistics only indicate statistical significance, not economic or practical significance of variables (Wooldridge, 2013, p. 135).

[14] Corruption elasticities in Sigman (2014) are computed by multiplying its estimated coefficients of corruption with its sample mean of corruption.

Table 9. | Tests for long-run causality.

	Weak exogeneity of BOD	Weak exogeneity of inequality
Model 1	6.91**	8.79**
Model 2	6.92**	7.31**
Model 3	10.39**	4.25**

**Rejection of the null hypothesis of weak exogeneity at the 5% level.

and corruption have different distributions, we also compute standardized coefficients for both variables. For column 1, we find that a one-standard deviation increase in inequality increases water pollution by 0.544 standard deviation; a one-standard deviation increase in corruption increases water pollution by 0.565 standard deviation. Again, the magnitude of inequality is nearly as large as that of corruption, meaning that inequality as almost important as corruption in determining water pollution.

Our results so far show that inequality increases water pollution. An increase in water pollution may also lead to an increase in inequality because bottom class people are more exposed to water pollution (and thus more vulnerable to waterborne diseases)[15]. Sickness reduces productivity and increases absenteeism at work and in school. This consequently reduces the earning capacity of bottom class adults and the potential income of their children, perpetuating poverty, and reproducing inequality. Therefore, the direction of causality may run in both directions – not only from inequality to water pollution but also from water pollution to inequality. To test for long-run causality, this paper performs a test of weak exogeneity (a test of zero restrictions in the α matrix). A rejection of the null of weak exogeneity indicates long-run (Granger) causality (Hall & Milne, 1994). In Table 9, the null hypothesis of weak exogeneity of water pollution and the null hypothesis of weak exogeneity of inequality are rejected at the 5% level; the long-run Granger causality runs in both directions, from inequality to water pollution and water pollution to inequality. These results even hold when income is included in our model (model 2) and when Gini is used instead of Theil (model 3). This finding justifies our use of the Johansen approach that views all variables to be endogenous.

The long-run causality tests indicate the direction of the relationship (i.e., water pollution to inequality), but do not determine the sign of that relationship. Does an increase in water pollution lead to an increase in inequality? To see this, we compute an impulse-response function from our estimated VECM. Figure 1 plots the response of inequality to a shock in water pollution for a period of 10 years. As a robustness check, impulse-response function using Gini is also reported. In both panels, we find a shock to water pollution does have a significant permanent positive effect on inequality[16]. For Theil, the full impact is reached after 2 years, and 5 years for Gini. Furthermore, the bootstrapped confidence interval in both panels lies almost entirely in the positive domain. We conclude that there is a positive effect of water pollution on inequality.

Future research that examines the effect of inequality on environmental outcomes may need to tackle endogeneity bias that arises from reverse causality going from more pollution to higher inequality. The reason is that reverse causality going from more pollution to higher inequality would bias the estimated effect of inequality on pollution upward. To put it another way, if pollution has a positive effect on inequality, then the effect of inequality on pollution tends to be positive.

---

[15] Access to clean water in India is correlated with a person's social class (Zargar, 2018).

[16] The effects of shocks can be either transitory or permanent (Gonzalo & Ng, 2001).
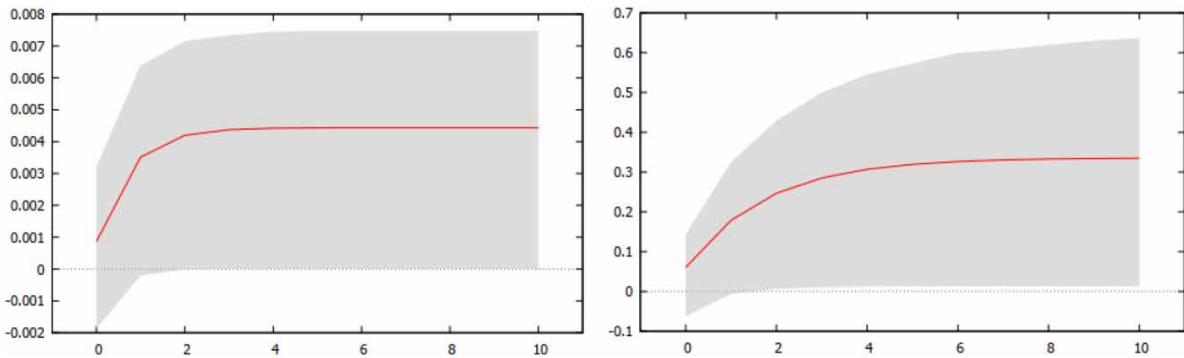
Fig. 1. | Impulse responses of Theil (left) and Gini (right) to a one-standard deviation shocks in water pollution. Bootstrap 95% confidence bounds are based on 2000 replications.

6. CONCLUSION

Although India is notorious for its water quality and inequality, little attention has been paid to the question of whether these two phenomena are related. Previous literature that examined the relationship between water quality and inequality did not explicitly analyze for the case of India. Although lessons can still be learned from the empirical literature, it suffered from many econometric issues, which were rectified in the present study.

We found that inequality has a robust positive long-run effect on water pollution in India. One reason why inequality increases water pollution in India could be that (in no particular order) the rich consume more water than is necessary and thus produce more pollution. Another reason could be that the wealthy set the norm for what defines a lifestyle of high status, forcing people to spend more and consume more (and more production in the economy as a result). Another plausible reason is that inequality always erodes trust, making cooperation to protect common resources become difficult. The next reason is that inequality causes people to think more about economic issues and shifts people's attention away from environmental problems, thereby reducing demand for better environment. The reader should bear in mind that these are just possible causal pathways, and more research is needed to develop a deeper understanding of the interplay between inequality and water pollution in India.

We showed that the magnitude of inequality on water pollution is almost as large as that of corruption, a factor that is often seen as an important cause of water pollution (Rowlatt, 2016). This suggests that reducing inequality is almost as important as curbing corruption in addressing water challenges in India. Given that an increase in water pollution leads to an increase in inequality, future studies may need to consider the possible endogeneity of inequality when assessing the impact of inequality on water pollution or other environmental outcomes.

7. POLICY RECOMMENDATION

From a policy point of view, reducing inequality is not only important for fairness but also for the environment. In 2016, India discontinued the wealth tax due to high collection cost, and in 2019, the government announced a corporate tax cut to spur the economy. This study recommends the government reconsider their decisions since inequality may be worsened as a result (Jones, 2015; Nallareddy *et al.*, 2018) and can thereby have an unintended effect on its water. Although reducing tax collection may spur more economic growth, which according to the EKC hypothesis, could later be good for the environment (by increasing public demand for environmental protection, for example), we did not find the EKC for water pollution in India.

It is estimated that the corporate tax revenue loss due to the recent tax cut would be about $20 billion (around 0.7% of GDP) (The Economist, 2019b, para. 3). India also suffers a significant loss due to tax avoidance. For example, in 2013, India's corporate tax revenue loss due to tax avoidance is estimated to be between $41.17 billion and $47.53 billion, between 2.34 and 2.70% of GDP (Cobham & Janský, 2018). It is well known that, in reducing water pollution, the water sector in the country faces a huge financial challenge to fund investments in capital-intensive modern wastewater collection and treatment (Narain, 2016). If India could increase the amount of tax revenue, it could be used to finance the necessary investments. The tax revenue could also be used to fund vital public services that could help reduce inequality such as education and healthcare for people from all walks of life in India. The government can also consider consumption-based tax to reduce consumption among higher consumers, in addition to increasing tax collection (Mohan, 2019).

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Agrawal, P. K. (1994). *Environment Protection and Pollution Control in the Ganga*. MD Publications Pvt. Ltd, New Delhi.

Atkinson, A. B. & Brandolini, A. (2001). Promise and pitfalls in the use of 'secondary' data-sets: income inequality in OECD countries as a case study. *Journal of Economic Literature 39*, 771–799.

Bari, P. (2018). *Experts Flush Out India's 'Sewage' Rivers. Hindustan Times*. Available at: https://www.hindustantimes.com/pune-news/experts-flush-out-india-s-sewage-rivers/story-7UXx4jQmvxbM2yPV0ar4YJ.html

Barua, A. & Hubacek, K. (2009). An empirical analysis of the environmental Kuznets curve for water pollution in India. *International Journal of Global Environmental Issues 9*(1–2), 50–68.

Berthe, A. & Elie, L. (2015). Mechanisms explaining the impact of economic inequality on environmental deterioration. *Ecological Economics 116*, 191–200.

Bonham, C. S. & Cohen, R. H. (2001). To aggregate, pool, or neither: testing the rational-expectations hypothesis using survey data. *Journal of Business & Economic Statistics 19*(3), 278–291.

Cheung, Y. W. & Lai, K. S. (1993). Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics 55*(3), 313–328.

Chitnis, P. (2017). *India's Big Job Creating Industry Is Dying a Slow Death. BloombergQuint*. Available at: https://www.bloombergquint.com/business/indias-big-job-creating-industry-is-dying-a-slow-death

Clement, M. & Meunie, A. (2010). Is inequality harmful for the environment? an empirical analysis applied to developing and transition countries. *Review of Social Economy 68*(4), 413–445.

Cobham, A. & Janský, P. (2018). Global distribution of revenue loss from corporate tax avoidance: re-estimation and country results. *Journal of International Development 30*(2), 206–232.

Cole, M. A. & Neumayer, E. (2006). The impact of poor health on total factor productivity. *The Journal of Development Studies 42*(6), 918–938.

Cushing, L., Morello-Frosch, R., Wander, M. & Pastor, M. (2015). The haves, the have-nots, and the health of everyone: the relationship between social inequality and environmental quality. *Annual Review of Public Health 36*, 193–209.

Damania, R., Desbureaux, S., Rodella, A. S., Russ, J. & Zaveri, E. (2019). *Quality Unknown: The Invisible Water Crisis*. The World Bank, Washington DC.

Deaton, A. (2013). What's wrong with inequality? *The Lancet 381*(9864), 363.

DeFrancis, M. P. (2011). *Economic Impacts of Inadequate Sanitation in India*. Water and Sanitation Program, Washington, DC.

Dembowski, H. (2001). *Taking the State to Court: Public Interest Litigation and the Public Sphere in Metropolitan India*. Oxford University Press, USA. https://www.bookdepository.com/Taking-State-Court-Hans-Dembowski/9780195653090.

Desbureaux, S., Damania, R., Rodella, A. S., Russ, J. & Zaveri, E. (2019). *The Impact of Water Quality on GDP Growth: Evidence From Around the World*. The World Bank, Washington DC.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366a), 427–431.

DNA India (2019). BIS study: Mumbai tops ranking for quality of tap water, Delhi lowest. *DNA India*. Available at: https://www.dnaindia.com/india/report-bis-study-mumbai-tops-ranking-for-quality-of-tap-water-delhi-lowest-2801970

Do, Q. T., Joshi, S. & Stolper, S. (2018). Can environmental policy reduce infant mortality? evidence from the ganga pollution cases. *Journal of Development Economics 133*, 306–325.

Dorling, D. (2017). *Is Inequality Bad for the Environment? The Guardian*. Available at: https://www.theguardian.com/inequality/2017/jul/04/is-inequality-bad-for-the-environment

Durlauf, S. N., Johnson, P. A. & Temple, J. R. (2009). The methods of growth econometrics. In: *Palgrave Handbook of Econometrics* (T. C. Mills & K. Patterson, eds). Palgrave Macmillan, London, pp. 1119–1179.

Eriksson, C. & Persson, J. (2003). Economic growth, inequality, democratization, and the environment. *Environmental and Resource Economics 25*(1), 1–16.

Everaert, G. (2011). Estimation and inference in time series with omitted I(1) variables. *Journal of Time Series Econometrics 2*(2), 1–26.

Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review 90*(4), 869–887.

Galbraith, J. K. (2009). Inequality, unemployment and growth: new measures for old controversies. *The Journal of Economic Inequality 7*(2), 189–206.

Gassebner, M., Lamla, M. J. & Sturm, J. E. (2011). Determinants of pollution: what do we really know? *Oxford Economic Papers 63*(3), 568–595.

Goldar, B. & Banerjee, N. (2004). Impact of informal regulation of pollution on water quality in rivers in India. *Journal of Environmental Management 73*(2), 117–130.

Gonzalo, J. & Ng, S. (2001). A systematic framework for analyzing the dynamic effects of permanent and transitory shocks. *Journal of Economic Dynamics and Control 25*(10), 1527–1546.

Grafton, R. Q. & Knowles, S. (2004). Social capital and national environmental performance: a cross-sectional analysis. *The Journal of Environment & Development 13*(4), 336–370.

Granger, C. W. J. & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics 2*, 111–120.

Greenstone, M. & Hanna, R. (2011). *Environmental Regulations, Air and Water Pollution, and Infant Mortality in India (NBER Working Paper No. 17210)*. National Bureau of Economic Research, Washington, DC.

Greenstone, M. & Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in India. *American Economic Review 104*(10), 3038–3072.

Grossman, G. M. & Krueger, A. B. (1995). Economic growth and the environment. *The Quarterly Journal of Economics 110*(2), 353–377.

Gujarati, D. N. & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill, New York.

Gupta, S., Davoodi, H. & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Economics of Governance 3*(1), 23–45.

Hall, S. G. & Milne, A. (1994). The relevance of P-star analysis to UK monetary policy. *The Economic Journal 104*(424), 597–604.

Hassler, U. & Kuzin, V. (2009). Cointegration analysis under measurement errors. In: *Measurement Error: Consequences, Applications and Solutions* (J. M. Binner, D. L. Edgerton & T. Elger, eds). Emerald Group Publishing Limited, Bingley.

Heerink, N., Mulatu, A. & Bulte, E. (2001). Income inequality and the environment: aggregation bias in environmental Kuznets curves. *Ecological Economics 38* (3), 359–367. https://doi.org/10.1016/S0921–8009(01)00171–9

Herzer, D. (2020). How does mortality affect innovative activity in the long run? *World Development 125*, 104688.

Hirani, P. & Dimble, V. (2019). *Water Pollution is Killing Millions of Indians. Here's How Technology and Reliable Data Can Change That*. Available at: https://www.weforum.org/agenda/2019/10/water-pollution-in-india-data-tech-solution/

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The Stata Journal 7*(3), 281–312.

Islam, S. N. (2015). *Inequality and Environmental Sustainability (DESA Working Paper No. 145)*. United Nations, New York.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregessive Models*. Oxford University Press, New York.

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica 70*(5), 1929–1961.

Jones, C. I. (2015). Pareto and Piketty: the macroeconomics of top income and wealth inequality. *Journal of Economic Perspectives 29*(1), 29–46.

Jun, Y., Zhong-kui, Y. & Peng-fei, S. (2011). Income distribution, human capital and environmental quality: empirical study in China. *Energy Procedia 5*, 1689–1696.

Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, New York.

Karn, S. K. & Harada, H. (2001). Surface water pollution in three urban territories of Nepal, India, and Bangladesh. *Environmental Management 28*(4), 483–496.

Kasuga, H. & Takaya, M. (2017). Does inequality affect environmental quality? evidence from major Japanese cities. *Journal of Cleaner Production 142*, 3689–3701.

Knowles, S. (2005). Inequality and economic growth: the empirical relationship reconsidered in the light of comparable data. *The Journal of Development Studies 41*(1), 135–159.

Krieger, T. & Meierrieks, D. (2019). Income inequality, redistribution and domestic terrorism. *World Development 116*, 125–136.

Lahiri, K. & Mamingi, N. (1995). Testing for cointegration: power versus frequency of observation – another view. *Economics Letters 49*(2), 121–124.

Lin, C. Y. C. & Liscow, Z. D. (2013). Endogeneity in the environmental Kuznets curve: an instrumental variables approach. *American Journal of Agricultural Economics 95*(2), 268–274.

Mani, M., Markandya, A., Sagar, A. & Strukova, E. (2012). *An Analysis of Physical and Monetary Losses of Environmental Health and Natural Resources in India (Policy Research Working Paper No. 6219)*. World Bank, Washington, DC.

Mohan, D. (2019). *What Should India's Tax Reform Trajectory Look Like? The Wire*. Available at: https://thewire.in/economy/consumption-based-tax-system-reform-india

Mohan, D. (2020). *Two Wheels Good: India Falls Back in Love With Bikes After COVID-19. The Guardian*. Available at: https://www.theguardian.com/global-development/2020/jul/24/two-wheels-better-than-four-india-falls-in-love-with-cycling-lockdown-coronavirus

Müller-Fürstenberger, G. & Wagner, M. (2007). Exploring the environmental Kuznets hypothesis: theoretical and econometric problems. *Ecological Economics 62*(3–4), 648–660.

Nallareddy, S., Rouen, E. & Suárez Serrato, J. C. (2018). *Do Corporate Tax Cuts Increase Income Inequality? (NBER Working Paper No. 24598)*. National Bureau of Economic Research, Washington, DC.

Narain, S. (2016). Poverty and environmental inequality in India. In: *World Social Science Report 2016: Challenging Inequalities: Pathways to a Just World*. ISSC, IDS, & UNESCO (ed.). UNESCO Publishing, Paris, pp. 137–139.

Narayan, P. K. & Narayan, S. (2010). Carbon dioxide emissions and economic growth: panel data evidence from developing countries. *Energy Policy 38*(1), 661–666.

Navarro, V., Muntaner, C., Borrell, C., Benach, J., Quiroga, Á., Rodríguez-Sanz, M., Vergés, N. & Pasarín, M. I. (2006). Politics and health outcomes. *The Lancet 368*(9540), 1033–1037.

Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics 1. *Oxford Bulletin of Economics and Statistics 54*(3), 461–472.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica: Journal of the Econometric Society 60*, 119–143.

Phillips, P. C. & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies 57*(1), 99–125.

Phillips, P. C. & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika 75*(2), 335–346.

Ravi Rajan, S. (2014). A history of environmental justice in India. *Environmental Justice 7*(5), 117–121.

Reinsel, G. C. & Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: estimation. likelihood ratio test, and forecasting. *Journal of Time Series Analysis 13*(4), 353–375.

Rowlatt, J. (2016). *India's Dying Mother. BBC News*. Available at: https://www.bbc.co.uk/news/resources/idt-aad46fca-734a-45f9-8721-61404cc12a39

Schmith, T., Johansen, S. & Thejll, P. (2012). Statistical analysis of global surface temperature and sea level using cointegration methods. *Journal of Climate 25*(22), 7822–7833.

Scruggs, L. A. (1998). Political and economic inequality and the environment. *Ecological Economics 26*(3), 259–275.

Seldadyo, H. & De Haan, J. (2011). Is corruption really persistent? *Pacific Economic Review 16*(2), 192–206.

Sigman, H. (2014). Decentralization and environmental quality: an international analysis of water pollution levels and variation. *Land Economics 90*(1), 114–130.

Stern, D. I. (2011). The role of energy in economic growth. *Annals of the New York Academy of Sciences 1219*, 26–51. https://doi.org/10.1111/j.1749–6632.2010.05921.x.

Stock, J. H. & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica: Journal of the Econometric Society 61*, 783–820.

Stock, J. H. & Watson, M. W. (2015). *Introduction to Econometrics*. Pearson Education Limited, Essex.

The Economist (2019a). The world's most sacred river – the Ganges – is also one of its dirtiest. *The Economist*. Available at: https://www.economist.com/asia/2019/03/28/the-worlds-most-sacred-river-the-ganges-is-also-one-of-its-dirtiest

The Economist (2019b). India's government delights businesses by slashing corporate tax. *The Economist*. Available at: https://www.economist.com/finance-and-economics/2019/09/26/indias-government-delights-businesses-by-slashing-corporate-tax

The Hindu (2017). Affluent people use four times more water than the average household. *The Hindu*. Available at: https://www.thehindu.com/news/cities/bangalore/affluent-people-use-four-times-more-water-than-the-average-household/article21938348.ece

Thompson, A. (2016). Downstream dumping. *Environmental Economics and Policy Studies 18*(3), 437–441.

Torras, M. & Boyce, J. K. (1998). Income, inequality, and pollution: a reassessment of the environmental Kuznets curve. *Ecological Economics 25*(2), 147–160.

Wilkinson, R. & Pickett, K. (2010). *The Spirit Level: Why Equality Is Better for Everyone*. Bloomsbury Press, New York.

Wisman, J. D. (2011). Inequality, social respectability, political power, and environmental devastation. *Journal of Economic Issues 45*(4), 877–900.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western, Mason.

Zargar, H. (2018). *India Has World's Highest Inhabitants Without Safe Water: Report. Mint*, p. 1. Available at: https://www.livemint.com/Politics/WoHowGWquof0lr7KPDV7GO/India-has-worlds-highest-inhabitants-without-safe-water-re.html