



Research article

Effect of environmental and socio-economic factors on the spreading of COVID-19 at 70 cities/provinces

Jishan Ahmed^{a,b}, Md. Hasnat Jaman^c, Goutam Saha^{d,*}, Pratyya Ghosh^d^a Department of Computer Science, Bowling Green State University, USA^b Department of Mathematics, University of Barisal, Bangladesh^c Department of Geology and Mining, University of Barisal, Bangladesh^d Department of Mathematics, University of Dhaka, Bangladesh

ARTICLE INFO

Keywords:

Temperatures
Humidity
Air quality
Population density
GDP
Health expenditure
Life expectancy
Total test
COVID-19

ABSTRACT

The main goal of this article is to demonstrate the impact of environmental and socio-economic factors on the spreading of COVID-19. In this research, data has been collected from 70 cities/provinces of different countries around the world that are affected by COVID-19. In this research, environmental data such as temperatures, humidity, air quality and population density and socio-economic data such as GDP (PPP) per capita, per capita health expenditure, life expectancy and total test in each of these cities/provinces are considered. This data has been analyzed using statistical models such as Poisson and negative binomial models. It is found that a negative binomial regression model is the best fit for our data. Our results reveal higher population density to be an important factor for the quick spread of COVID-19 as maintenance of social distancing requirements are more difficult in urban areas. Moreover, GDP (PPP) and PM_{2.5} are linked with fewer cases of COVID-19 whereas PM₁₀, and total number of tests are strongly associated with the increase of COVID-19 case counts.

1. Introduction

In December 2019, a new RNA virus strain from the family *Coronaviridae* emerged in Wuhan, the capital of Hubei province (Wu et al., 2020). This novel virus is a betacoronavirus and designated as SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) causing a pneumonia disease called coronavirus disease 2019 (COVID-19) (Gorbalenya et al., 2020). Though SARS-CoV-2 has a low mortality rate (about 2.3%) compared to other coronaviruses like SARS-CoV (about 10%) and MERS-CoV (about 35%), the reproduction number or transmission rate of SARS-CoV-2 virus is very high (2.24–3.58) (Ceccarelli et al., 2020; Zhao et al., 2020) causing rapid spreading and becoming a pandemic. Though fever, fatigue and dry cough are the most common symptoms, some patients can develop severe and even fatal complications such as Acute Respiratory Distress Syndrome (ARDS) (Wang et al., 2020a, 2020b).

Coronaviruses are enveloped viruses which predominantly deputize through outright contiguity with respiratory droplets of an infected person (generated through coughing and sneezing). By touching one's own face (i.e. eyes, ears, nose, and mouth) after touching a surface contaminated with the virus a distinct person can also be infected.

Enveloped viruses can survive for several hours on different surfaces, however they show sensitivity to heat, detergent and desiccation compared to non-enveloped viruses (Howie et al., 2008). Therefore, environmental factors may have a great impact on the transmission of infectious disease by affecting the survival of coronavirus on surfaces or in the air (Casanova et al., 2010). High temperature and high relative humidity environments reduced the transmission of SARS coronavirus (Chan et al., 2011). Ma et al. (2020) found that both 1 unit increase of temperature and absolute humidity were related to a decrease in COVID-19 deaths. Some other studies (Hongchao et al., 2020; Oliveiros et al., 2020; Tosepu et al., 2020; Wang et al., 2020a, 2020b) also support a relation between environmental factors and COVID-19 such as spread decreases with increase in temperature. Along with these factors population density and mobility can trigger the spreading of this virus (Oztig and Askin, 2020).

The main goal of this research is to provide scientific evidence based on statistical modeling regarding the spreading of the SARS-CoV-2 under the changing circumstances of humidity, temperatures, population density, GDP (PPP), life expectancy, health expenditure, total tests, air quality index, and particulate matters such as PM_{2.5} and PM₁₀.

* Corresponding author.

E-mail address: gsahamath@du.ac.bd (G. Saha).

2. Materials and methods

In this research, data containing the total number of infected cases, death count by COVID-19, population density, GDP (PPP) per capita, per capita health expenditure, life expectancy, total test, monthly average humidity, average high and low temperatures have been collected from 126 cities/provinces of 42 most affected countries around the world from January 18, 2020 to April 24, 2020. Weather data was collected from AccuWeather (<http://www.accuweather.com/>) in order to get reliable data. Such reliable data is important in order to get the correct and accurate research findings to understand the impact of weather on spreading the COVID-19. Also, population data like area, population density were collected from worldometer (www.worldometers.info/population). In addition, socio-economic data like GDP (PPP) per capita in 2017, life expectancy of people of different country and total number of COVID-19 test are collected from worldometer (<http://www.worldometers.info/population>) and data for per capita health expenditure in 2017 is collected from Knoema (<https://knoema.com/atlas>). Also, air quality index, and particulate matters such as PM_{2.5}, PM₁₀ data are collected from plume labs (<https://air.plumelabs.com/en/>).

USA, France, and a few other countries counted suspected COVID-19 cases as a COVID-19 death to include uncounted fatalities due to the lack of massive testing capacity. Belgium even included flu-like symptomatic deaths as COVID-19 death. Countries are still struggling to scale the testing capacity. Due to this limitation, we verified our data using multiple sources like google coronavirus (COVID-19) statistics data, worldometer (www.worldometers.info/coronavirus) and COVID-19 related pages of government website for different countries. These sources are presented in the appendix.

We had missing values of death count and infected cases for some cities. We really don't know why they were missing. In our study, we were careful to make our recommendations and refrained making predictions due to this scarcity of the data. Since the presence of missing values in the data can reduce the statistical power of a study, it can lead us to invalid conclusions along with biased estimates. Therefore, we have considered the case deletion method, the most common approach to

handle the missing data instead of applying other imputation techniques like maximum likelihood method. After handling missing values, we ended up with 24 countries and 70 cities/provinces. We performed our statistical analysis using these cities/provinces. Our implementations can be reproduced using the code and data made available at our GitLab repository <https://gitlab.com/Jishan/covid-19-research-2020.git>.

Two different models such as negative Binomial, and Poisson models are considered in this research. These models are assessed using the Akaike Information Criterion (AIC), Cragg & Uhler's (Oztig and Askin, 2020) pseudo-R², residual deviance, and Pearson statistic. We have used glm() function to fit the Poisson and glm.nb() function to fit the negative Binomial model within the R software (<https://cran.r-project.org/src/base/R-4>) package MASS (R Core Team, 2017). Two-sided statistical tests were considered along with 5% significance level.

2.1. Data observation

Figure 1 shows the distribution of the number of infected people in 70 cities/provinces. Our initial observations suggest that there could be a relationship between the environmental parameters and expansion of COVID-19 across the different geographical locations. Most of the cities/provinces where outbreaks occurred such as Madrid, New York etc., had low temperature and/or low to moderate humidity probably because coronavirus can survive longer on surfaces or respiratory droplets at this environmental condition. Places with relatively high humidity and high temperature such as Banten, Central Luzon etc., showed comparatively less infected people. Another factor, population density and mobility alone can trigger the infection rate logarithmically irrespective of environmental condition. São Paulo, Riyadh etc. cities had high temperature and high humidity but many infected people due to population density and mobility. In cold regions, population density can exacerbate the total COVID-19 infection along with the environment. According to the data analysis, our observation illustrates that there could be a remarkable connection between the environmental, socio-economic parameters and

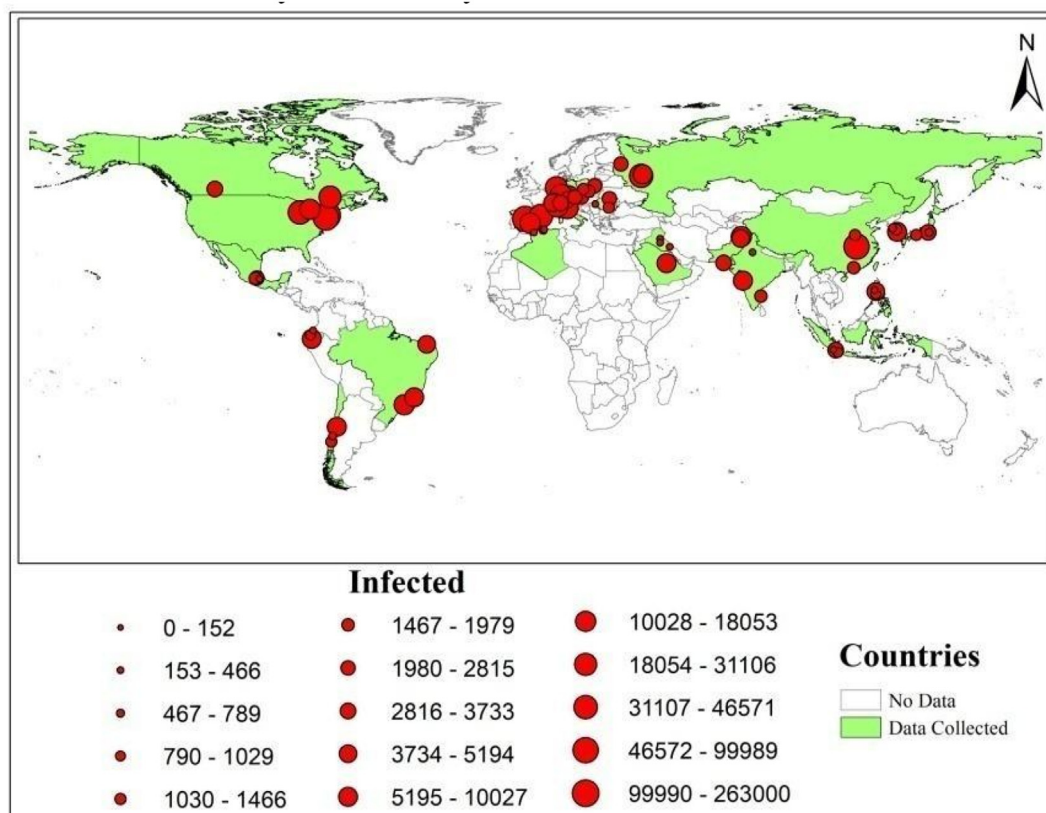


Figure 1. Distribution of COVID-19 Infected population.

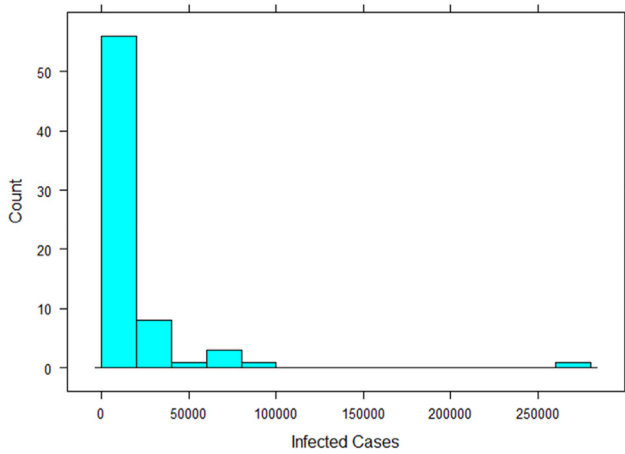


Figure 2. Distribution of Infected case.

the nature of the COVID-19 virus. In the next section, we will present statistical analysis and try to understand the above mentioned behavior.

2.2. Statistical analysis

In this paper, a generalized linear model (GLM) framework (Agresti, 2015) for count data has been deployed to analyze the effects of population density, GDP (PPP), health expenditure, life expectancy, total COVID-19 tests, humidity, temperatures and air pollutants on the spreading of COVID-19.

In the Poisson regression model, it is assumed that the variance and mean of the dependent variable are the same. However, this assumption is not always true, especially while studying the environmental risk to human health due to the fact that the variance is higher than average causing the overdispersion of the data. It is challenging to handle overdispersion in the modeling of count response variables like the number of COVID-19 confirmed cases. In our data, the variance of the infected cases is 1,279,339,997 and mean is 14,924.97 - variance is larger than the mean. Also, from Figure 2, we see that our response variable, the count of infected cases is highly skewed. This indicates that our data may be overdispersed. It is convenient to use a negative binomial model to estimate the parameter due to the presence of overdispersion of the data. Therefore, in this study, we have considered the negative binomial model and compared our results with the Poisson model as well to detect overdispersion in our data.

3. Results

3.1. Dataset descriptive analysis

In this work, we considered 70 cities/provinces around the world that had the confirmed cases of COVID-19. Figure 3 shows all explanatory variables through normalized heat map representations. The color scale on the right represents the intensity of the variables according to the saturation level of this scale. For example, New York had the highest number of COVID-19 confirmed cases which is shown in this figure with a highly saturated blue color.

Now, Pearson correlation coefficients are computed as shown in Figure 4 to determine the possible effects of collinearity. It shows that the Pearson

correlation between explanatory variables along with the significance measure. There was a strong positive correlation between average high and low temperatures. GDP (PPP), and health expenditure were positively correlated. It was also noticeable that air pollutants PM₁₀, and PM_{2.5} had positive correlation as well. However, the presence of high correlation among predictor variables does not violate any assumptions of GLMs.

In the following, we presented the summary statistics for infected cases, population density, GDP (PPP) per capita, per capita health expenditure, life expectancy, total test, humidity and average high and low temperatures as shown in Table 1. It is to be noted that the average number of confirmed infected cases was 14,925, the mean value of population density was 4,043.4 per km², the mean values of humidity, average high and low temperatures were 65.28%, 20.42 °C and 9.41 °C.

3.2. Selection and adjustment of the regression models

Since the dependent variable, COVID-19 case count is highly-skewed and non-continuous, standard linear regression models such as ordinary least squares regression are not appropriate for this count data. Therefore, the Poisson log-linear model would be our first-choice modeling technique. The expected COVID-19 infected case count (μ_i parameter) in the Poisson log-linear model is estimated as

$$\log(\mu_i) = \sum_{j=1}^p \beta_j x_{ij} \text{ or } \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \tag{1}$$

where, $\boldsymbol{\beta}$ is a vector of estimated coefficients of exploratory variables including the logarithm of the population density, GDP (PPP) per capita, per capita health expenditure, life expectancy and total test along with humidity, average high temperature, average low temperature, AQI, PM_{2.5}, and PM₁₀. For the sake of simplicity, we referred population density as PopDensity, GDP (PPP) per capita as GDPPPP, per capita health expenditure as HealthExpend, average high temperature as AvgHigh, average low temperature as AvgLow, PM₁₀ as PM1, PM_{2.5} as PM2 during the model building.

It is assumed in Poisson distribution that the mean and the variance are equal to the μ_i parameter. However, this assumption was not satisfied for the data used in this study. The greater ratio of variance to mean leads to overdispersion. The problem of overdispersion is evident from the Poisson model fit as the ratio of residual deviance and degrees of freedom is 9311.293 which is greater than the dispersion parameter limit 1. Pearson statistic and the deviance statistic were used as well to assess the overall performance of the fitted model. In Table 2a, the p-values suggest that the Poisson model is not adequate suggesting a poor fit. We cannot even trust the p values due to the presence of substantial overdispersion in our Poisson model. Since the negative-binomial (NB) model is a different generalization of the Poisson that allows for over-dispersion, we apply the NB model to overcome this problem of over dispersion. A gamma-distributed error term (Oztig and Askin, 2020) is included to Eq. (1) to relax the Poisson model assumption by introducing additional randomness as

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \tag{2}$$

where ε_i follows a gamma distribution with mean 1 and variance α .

The NB model has a mean μ_i and variance $\mu_i + \alpha\mu_i^2$, where α is the overdispersion parameter which is used as a measure of dispersion. Therefore, we have considered the following NB regression model,

$$\begin{aligned} \text{Log}(\mu_i) = & \beta_0 + \beta_1 \log(1 + \text{PopDensity}) + \beta_2 \log(1 + \text{GDPPPP}) + \beta_3 \text{LifeExpectancy} + \beta_4 \log(\text{HealthExpend}) + \beta_5 \text{Humidity} + \beta_6 \text{AvgHigh} \\ & + \beta_7 \text{AvgLow} + \beta_8 \text{AQI} + \beta_9 \text{PM1} + \beta_{10} \text{PM2} + \beta_{11} \log(\text{TotalTest}) \end{aligned}$$

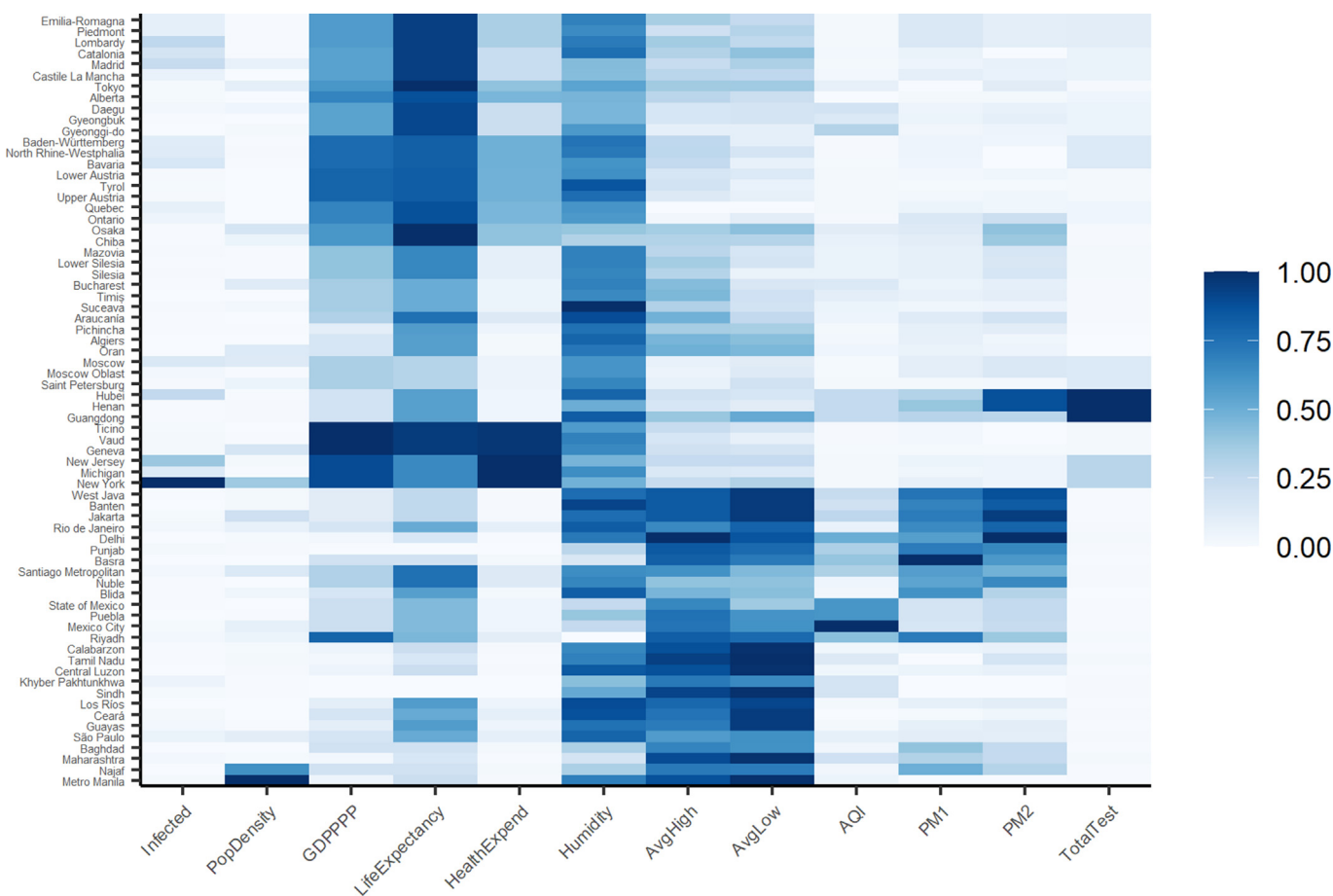


Figure 3. Normalized explanatory variables by cities/provinces.

In the NB model, we have found that the ratio of residual deviance and degrees of freedom is approximately equal to the dispersion parameter limit 1. The high Pseudo-R² values in Table 2b clearly indicate no evidence of lack-of-fit. It is to be mentioned here that the AIC score for the NB model was 1373.1.

3.3. Negative binomial model assessment

Diagnostic plots were used to check the performance of the negative Binomial model. In Figure 5, the jackknife deviance residuals vs. the fitted values are displayed on the top left panel, and normal QQ plots of the standardized deviance residuals are shown on the top right panel. The dotted line of the normal QQ plots represents the expected line if the standardized residuals are normally distributed (Davison and Snell, 1991). Cook statistics are shown in the bottom two panels. The bottom left plot shows the Cook statistics vs. the standardized leverages. The horizontal line is drawn at $8/(n-2p)$, and the vertical line is drawn at $2p/(n-2p)$, where n represents the number of observations and p represents the number of estimated parameters. Points above the horizontal line may be points which have high influence on the model. On the other hand, high leverage points correspond to the right side of the vertical line. We had 70 cities with COVID-19 confirmed cases, and 11 parameters were estimated. Here, we get a pretty accurate picture from the Figure 5 that our model is adequately describing the over dispersion in the count data when we use the negative binomial regression, but we may have some issues with extreme data points. Since the deletion of extreme data points may cause the other problems of over-fitting, it is not convenient to delete the outliers to increase the goodness-of-fit and power of explanation. However, we investigated these extreme data points in Figures 6 and 7 to detect the influential observations using the

Cook distance. In Figure 6, we see from the influence plot that observations 30 (Jakarta, Indonesia), and 45 (Khyber Pakhtunkhwa, Pakistan) stand out with large positive residuals whereas observations 16 (Guangdong, China) have large negative residuals. Observations 42 (Mexico City, Mexico), and 58 (Riyadh, Saudi Arabia) have a large leverage. However, it is evident from the residuals vs leverage plot (Figure 7) that none of them are influential observations. We evaluated our model without these outliers as well to see their impact. We found that outliers had no impact on the model performance as well. It does make sense because Jakarta had really extreme values of AQI, PM_{2.5}, and PM₁₀, Mexico had high values of AQI, PM_{2.5}, and PM₁₀ values were not available for Khyber Pakhtunkhwa, and Riyadh had low humidity with high values of AQI, PM_{2.5}, and PM₁₀.

Table 3 shows the associations of GDP (PPP) per capita, life expectancy, per capita health expenditure, total test, population density, humidity, average high temperature, average low temperature, AQI, PM_{2.5}, PM₁₀ with COVID-19 infected incidence. The results show that population density (Coefficient estimate: 0.135; 95% CI: (0.019, 0.255), p-value = 0.021), and GDP (Coefficient estimate: -1.631; 95% CI: (-2.931, -0.375), p-value = 0.021), PM₁₀ (Coefficient estimate: 0.017; 95% CI: (0.002, 0.033), p-value = .011), PM_{2.5} (Coefficient estimate: -0.022; 95% CI: (-0.037, -0.006), p-value = 0.001), total test (Coefficient estimate: 0.809; 95% CI: (0.610, 1.008), p-value = 0.000) were significantly associated with COVID-19. The results indicate that the “baseline” average, infected case count is 8142.499. We can interpret the other exponentiated coefficients multiplicatively as well.

Our results clearly demonstrate that for every unit increase in GDP, we estimated a significant decrease in COVID-19 infected case count of 80.4%. There is evidence to suggest that the percent change in the incident rate of COVID-19 infected case count is a 2.2% decrease for

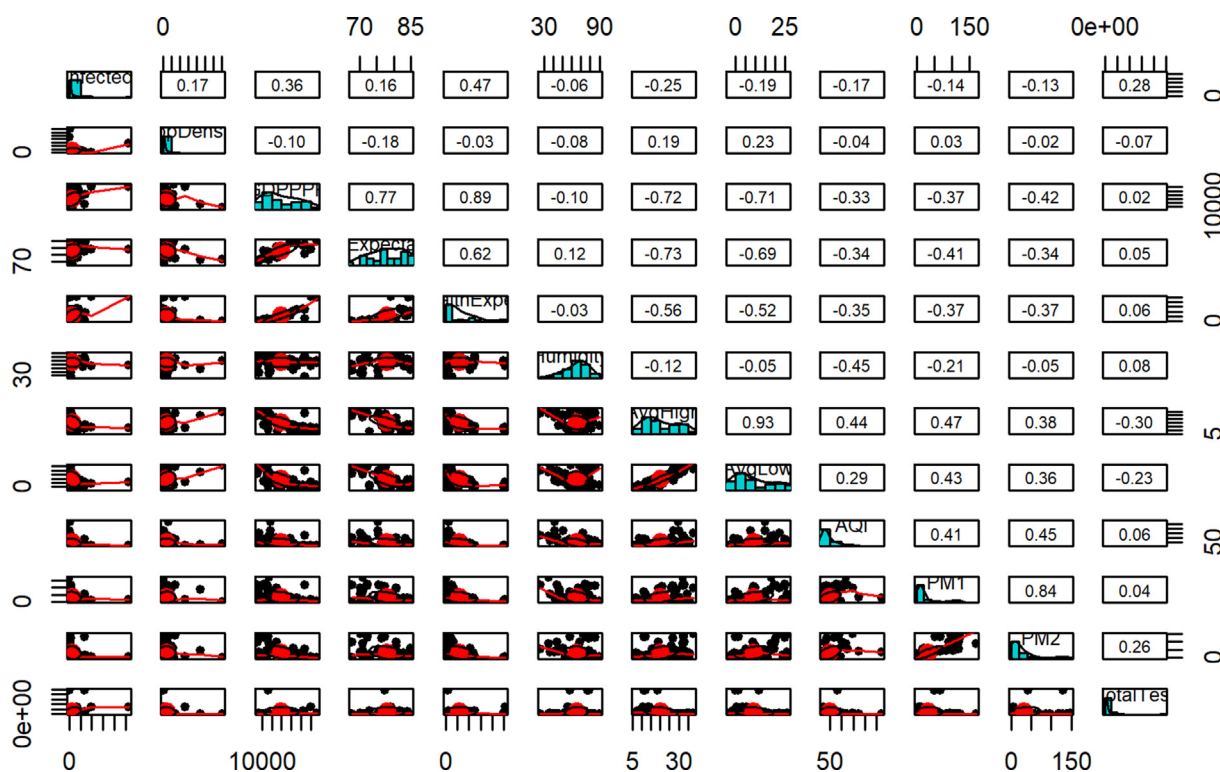


Figure 4. The pair wise plot along with Pearson correlation coefficients of infected cases, population density, GDP (PPP) per capita, life expectancy, per capita health expenditure, and total test along with humidity, average high temperature, average low temperature, AQI, PM_{2.5}, and PM₁₀.

every unit increase in PM_{2.5}. However, for every unit increase in population density, we could expect to see a 14.5% rise in COVID-19 infected case count. It is noticeable that each unit increase in PM₁₀ multiplies the COVID-19 infected case count by 1.017, a 1.7% increase. It is to be mentioned that the number of tests played a significant role to rise in the infected cases. It is even clearer from our result that for every unit increase in total test, we estimated a significant increase in COVID-19 infected case count of 124.6%.

4. Discussion

In this study, we attempted to answer the question of why some cities/provinces have higher numbers of COVID-19 infected people compared with others. In this study, we found from Table 3 that the population density, and GDP, PM₁₀, PM_{2.5}, and total tests were significantly associated with the COVID-19 confirmed infected cases.

Since COVID-19 is a highly contagious virus, population density can contribute to the spread of this virus (Coşkun et al., 2020). Oztig and

Askin (2020) examined the link between human mobility and the number of COVID-19 infected people in countries. They reported that countries that have higher population density (IRR = 2.403, p < 0.01) are found to be more likely to have higher numbers of COVID-19 infected cases than other countries. We also found statistically significant evidence that an increase of 1 unit in population density is associated with an 14.5% increase in the COVID-19 infected case count. It is difficult to maintain social distance in densely populated metropolitan cities and countries with tourist attractions. New York, New Jersey, Lombardy, Hubei, Madrid and Catalonia were the epicenter of the COVID-19 due to their dense population. New York, Lombardy, Madrid and Catalonia are popular tourist destinations as every year millions of tourists visit these cities. Taking the number of tourists into account when modeling the association between population density and COVID-19 could substantially improve the performance of our models. However, we did not consider the number of tourists as an explanatory variable due to the lack of reliable data.

Our results also indicate a positive association between the PM₁₀ and high numbers of COVID-19infected patients. It was surprising to see that countries with high values of air pollutants PM_{2.5} are less likely to have more COVID-19 cases than other countries. Associations between short-

Table 1. Summary statistics for different parameters.

| Variables | Mean | SD | Min | Max |
|--|--------|----------|--------|---------|
| Infected cases | 14925 | 35767.86 | 181 | 263000 |
| Population density (/km ²) | 4043.4 | 10563.71 | 6.0 | 71263.0 |
| GDP (PPP) per capita | 29715 | 17902.62 | 5539 | 66307 |
| Life expectancy | 77.98 | 5.01 | 67.79 | 85.03 |
| Per capita health expenditure | 2322.7 | 2917.15 | 45.0 | 10224.0 |
| Humidity (%) | 65.28 | 12.92 | 26.00 | 90.00 |
| Avg high temperature (°C) | 20.42 | 9.05 | 6.00 | 38.09 |
| Avg low temperature (°C) | 9.414 | 9.48 | -3.230 | 26.25 |
| AQI | 50.4 | 45.47 | 19.0 | 274.0 |
| PM _{2.5} | 32.17 | 40.18 | 0.00 | 148 |
| PM ₁₀ | 31.92 | 40.66 | 0.00 | 170 |
| Total test | 815174 | 2021204 | 3359 | 9800000 |

Table 2a. Goodness-of-fit (GOF) results for the Poisson model.

| Test | Value | df | p-value |
|----------|----------|----|---------|
| Deviance | 540055.4 | 58 | 0 |
| Pearson | 774514.4 | 58 | 0 |

Table 2b. Goodness-of-fit (GOF) result for the negative-binomial (NB) model.

| Metric | Value |
|-----------------------|-------|
| Pseudo-R ² | 0.858 |

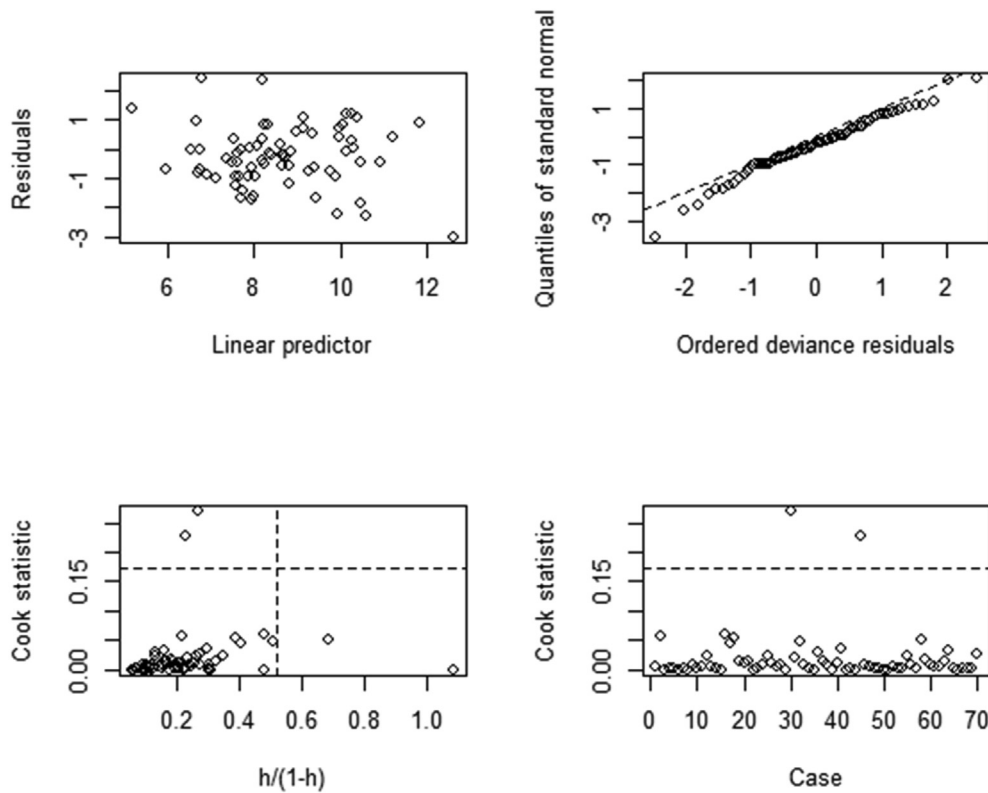


Figure 5. Diagnostic plots for the negative Binomial model.

term $PM_{2.5}$ exposure and poor infectious disease outcomes for influenza, pneumonia, and acute lower respiratory infections were reported by several previous studies (Croft et al., 2020; Horne et al., 2018). Therefore, we were expecting to see the positive association of $PM_{2.5}$ with COVID-19 like previous studies. Unfortunately, our study design cannot provide clear insight into the mechanisms underlying the negative relationship between $PM_{2.5}$ and COVID-19 infected case counts. We think that socio-economic indicators played a significant role here. For example, Jakarta, Indonesia had $140 \mu\text{g}/\text{m}^3$ for $PM_{2.5}$. In contrast, they have conducted only 27,075 tests due to the lack of testing equipment. We have seen from our results that the numbers of tests administered to individuals are significantly associated with the increased number of COVID-19 infected cases.

To stop the spread of the COVID-19, it is required to test more people and enable contact tracing. It is also equally important to incorporate mitigation techniques such as “Stay at Home” orders, frequent hand washing, and social distancing. However, it is quite impossible to implement such mitigation techniques for a country like India due to the widespread poverty and unequal distribution of income compared to a wealthy nation like Switzerland. Thus, we have attempted to explore the relationships between COVID-19 incidence with the socio-economic indicators such as GDP (PPP) per capita, life expectancy, and per capita health expenditure. We have found statistically significant evidence that GDP is associated with a decrease in the COVID-19 incident rate. It makes sense as people living in countries with higher GDP are likely to attend a

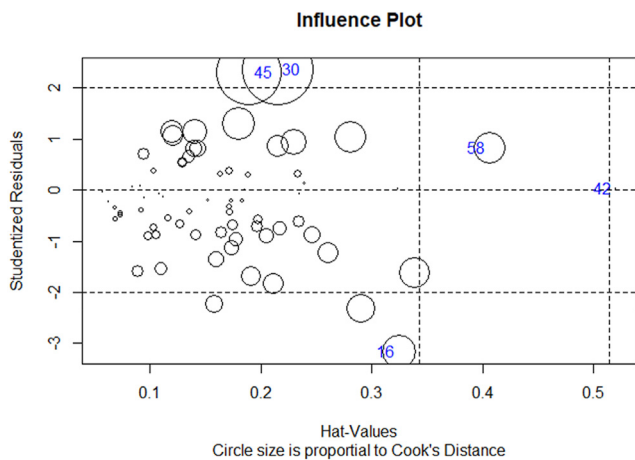


Figure 6. Influence plot for the negative binomial model.

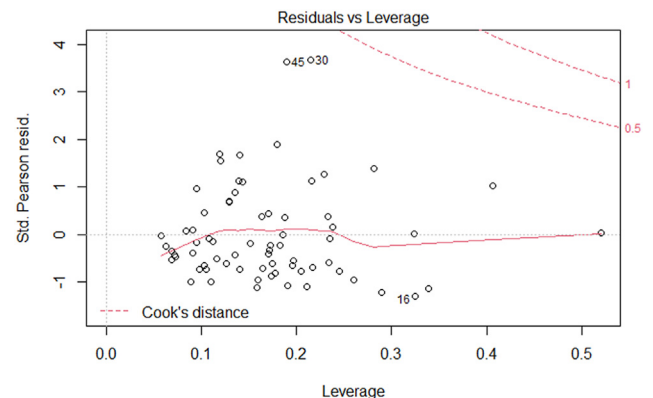


Figure 7. Residuals vs Leverage plot for the negative Binomial model.

Table 3. Estimation results of negative binomial regression.

| Effect | Estimate | Std. Error | 95% CI | IRR | p-value |
|------------------------|----------|------------|------------------|----------|---------|
| (Intercept) | 9.005 | 6.901 | (−3.055, 21.307) | 8142.499 | 0.197 |
| log(PopDensity) | 0.135 | 0.057 | (0.019, 0.255) | 1.145 | 0.021 |
| log(GDPPPP) | −1.631 | 0.685 | (−2.931, −0.375) | 0.196 | 0.021 |
| LifeExpectancy | 0.024 | 0.057 | (−0.087, 0.136) | 1.024 | 0.676 |
| log(1 + HealthExpend) | 0.610 | 0.326 | (0.015, 1.226) | 1.840 | 0.067 |
| Humidity | 0.003 | 0.011 | (−0.023, 0.028) | 1.003 | 0.767 |
| AvgHigh | −0.063 | 0.047 | (−0.165, 0.038) | 0.939 | 0.184 |
| AvgLow | 0.062 | 0.041 | (−0.020, 0.144) | 1.064 | 0.136 |
| AQI | 0.003 | 0.004 | (−0.005, 0.013) | 1.003 | 0.431 |
| PM10 | 0.017 | 0.006 | (0.002, 0.033) | 1.017 | 0.011 |
| PM2.5 | −0.022 | 0.006 | (−0.037, −0.006) | 0.978 | 0.001 |
| log(TotalTest) | 0.809 | 0.091 | (0.610, 1.008) | 2.246 | 0.000 |

The significance of bold in the table refers to the variables with $P < 0.05$.

larger number of social events and to spend more time travelling foreign countries possibly paving the way for easier virus diffusion. Also, the higher efficiency of national health systems allow them to administer more tests and that could affect the number of COVID-19 confirmed cases.

We observed fewer COVID-19 cases in warmer cities like Delhi and Mecca. Seasonal flu epidemics usually occur yearly during the colder months. COVID-19 is primarily spread from person to person through close contact. We can become infected from respiratory droplets when an infected person coughs, sneezes, or talks. Therefore, seasonal flu symptoms such as coughs and sneeze may contribute to the spread of the COVID-19 virus in the colder months (Mandal and Panwar, 2020). However, we did not find any evidence of association between COVID-19 cases and temperature. We believe that the proper mitigation techniques like flu vaccine alone could lower the COVID-19 cases during the flu seasons by limiting respiratory droplets. Our results were not convincing enough to infer that areas with average high temperatures are less likely to see the surge of massive COVID-19 cases. Also, average low temperature could drive the spread of virus through respiratory droplets. However, our model shows there is very little or no role of humidity in the outbreak of COVID-19. These findings are aligned with the review study conducted by Mecenas et al. (2020).

Since COVID-19 vaccines or effective drugs are still under development, identifying the environmental and socio-economic factors that intensify the spread of this virus would be helpful to design a better strategy to lower the spread as well for a future pandemic. Moreover, people from developing countries like Bangladesh may have to wait for two to three years to get the vaccine due to the tremendous demand of the vaccine. Already, Germany and the USA have signed a resolution that frontline health care workers will be vaccinated first. Motivated by this fact, we attempted to find the environmental and socio-economic factors that could intensify the spread of the COVID-19 virus in our study.

5. Conclusion

This work is by far the first attempt that selects a model by comparing two statistical models to understand the spreading of COVID-19. We found that the negative binomial provides the best fit to the data compared to the Poisson model. Our model infers that temperatures and humidity did not show any significant effect on the spreading of the virus. But population density has performed an extensive role for the spreading of COVID-19 in different countries. In addition, cities with higher population density pose extreme risk, which provides useful guidelines for policymakers and the public to control the COVID-19 pandemic. Most importantly, our research showed that GDP and PM_{2.5}

has a positive effect on the slowdown of spreading of COVID-19 infection whereas PM₁₀ and total tests significantly contributed to the rise of COVID-19 infection.

Declarations

Author contribution statement

J. Ahmed: Performed the experiments; Analysed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Md. Hasnat Jaman: Contributed reagents, materials, analysis tools or data; Wrote the paper.

G. Saha: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

P. Ghosh: Contributed reagents, materials, analysis tools or data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data associated with this study has been deposited at <https://gitlab.com/Jishan/covid-19-research-2020.git>.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2021.e06979>.

References

- Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. Wiley, Hoboken.
- Casanova, L.M., Jeon, S., Rutala, W.A., Weber, D.J., Sobsey, M.D., 2010. Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Appl. Environ. Microbiol.* 76, 2712–2717.
- Ceccarelli, M., Berretta, M., Rullo, E.V., Nunnari, G., Cacopardo, B., 2020. Differences and similarities between Severe Acute Respiratory Syndrome (SARS)-CoronaVirus (CoV) and SARS-CoV-2. Would a rose by another name smell as sweet? *Eur. Rev. Med. Pharmacol. Sci.* 23, 7–9.

- Chan, K.H., Peiris, J.S.M., Lam, S.Y., Poon, L.L.M., Yuen, K.Y., Seto, W.H., 2011. The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* 2011.
- Coşkun, H., Yıldırım, N., Gündüz, S., 2020. The spread of COVID-19 virus through population density and wind in Turkey cities. *Sci. Total Environ.* 751.
- Croft, D.P., Zhang, W., Lin, S., et al., 2020. Associations between source-specific particulate matter and respiratory infections in New York state adults. *Environ. Sci. Technol.* 54 (2), 975–984 [published Online First: 2019/11/23].
- Davison, A.C., Snell, E.J., 1991. Residuals and diagnostics. In: Reid, N., Snell, E.J. (Eds.), *Statistical Theory and Modelling: in Honour of Sir David Cox D.V. Hinkley*. Chapman and Hall, pp. 83–106.
- Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., Perlman, S., Poon, L.L.M., Samborskiy, D.V., Sidorov, I.A., Sola, I., Ziebuhr, J., 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544.
- Hongchao, Q., Shuang, X., Runye, S., P. W.M., Yue, C., Wei, T., Qing, S., Wenge, W., Xinyi, W., Zhijie, Z., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. *Sci. Total Environ.* 183135.
- Horne, B.D., Joy, E.A., Hofmann, M.G., et al., 2018. Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection. *Am. J. Respir. Crit. Care Med.* 198 (6), 759–766 [published Online First: 2018/04/14].
- Howie, R., Alfa, M.J., Coombs, K., 2008. Survival of enveloped and non-enveloped viruses on surfaces compared with other micro-organisms and impact of suboptimal disinfectant exposure. *J. Hosp. Infect.* 69, 368–376.
- Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., Luo, B., 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* 724, 138226.
- Mandal, C.C., Panwar, M.S., 2020. Can the summer temperatures reduce COVID-19 cases? *Publ. Health* 185, 72–79.
- Mecenas, P., Bastos, R.T.d.R.M., Vallinoto, A.C.R., Normando, D., 2020. Effects of temperature and humidity on the spread of COVID-19: a systematic review. *PLoS One* 15 (9), e0238339, 1–21.
- Oliveiros, B., Caramelo, L., Ferreira, N.C., Caramelo, F., 2020. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. *medRxiv*.
- Oztig, L.I., Askin, O.E., 2020. Human mobility and coronavirus disease 2019 (COVID-19): a negative binomial regression analysis. *Publ. Health* 185, 364–367.
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Tosepu, R., Gunawan, J., Effendy, D.S., Ahmad, L.O.A.I., Lestari, H., Bahar, H., Asfian, P., 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci. Total Environ.* 725, 138436.
- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., Peng, Z., 2020a. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA, J. Am. Med. Assoc.* 323, 1061–1069.
- Wang, J., Tang, K., Feng, K., Lv, W., 2020b. High temperature and high humidity reduce the transmission of COVID-19. *SSRN Electron. J.*
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269.
- Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., Wang, M.H., 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* 92, 214–217.