RESEARCH ARTICLE

WILEY

The density paradox: Are densely-populated regions more vulnerable to Covid-19?

Imad A. Moosa[1] | Ibrahim N. Khatatbeh[2]

[1]RMIT University, Melbourne, Australia

[2]The Hashemite University, Zarqa, Jordan

Correspondence
Ibrahim N. Khatatbeh, Department of Banking and Financial Sciences, Faculty of Economics and Administrative Sciences, The Hashemite University, Zarqa, Jordan.
Email: ibrahim.khatatbeh@hu.edu.jo

Abstract

The "density paradox" refers to the observation that some highly populated cities and countries have recorded a smaller number of Covid-19 cases than regions that are sparsely populated. We present empirical evidence on the role played by population density in spreading the coronavirus, based on cross-sectional data covering 172 countries (obtained from several sources, including the European Centre for Disease Prevention and Control, the World Bank and the Center for Health Security). The results, obtained by using the techniques of extreme bounds analysis (EBA) and variable addition tests, show that population density has a significantly positive effect on the number of cases but not the number of deaths, as the latter is better explained by measures of preparedness. Plausible explanations are presented for the results to conclude that the "density paradox" is not really a paradox. This paper makes a contribution by shedding more light on a frequently debated issue by using a completely different, and more robust, statistical techniques and by providing results that may be helpful for health and urban planners.

KEYWORDS
covid-19, population density, global health security index, extreme bounds analysis, variable addition tests

1 | INTRODUCTION

The rise of Covid-19 has led to the emergence of several controversies about the pandemic and how to deal with it, including some conspiracy theories on its origin. More serious issues than conspiracy theories are controversies about, amongst others, the effectiveness of social distancing, the pros and cons of lockdown, and the choice between putting lives or the economy first (even though the loss of life could come from the pandemic or a collapsing economy). This paper is about another controversy, the role played by population density in the spread of the virus.

Views on the effect of population density on the spread of contagious disease (Covid-19 and otherwise) range between "unimportant" and "important". In between, we find views such as "important but less so than other factors" and "important but not as important as it is portrayed to be". There are also those who think that the evidence is inconclusive. Supportive evidence is produced by, among others, Maybery,[1] Sumdani et al.[2] Kadi and Khelfaoui[3] and Carozzi et al.[4] Wong[5] argues that how far people can be spatially separated is partly behavioural but partly constrained by population density. Johnston[6] notes that "population density is playing a role in transmission rates but it's not the only factor", suggesting that "the stats so far already muddle this correlation, with densely populated places such as Hong Kong and Singapore managing the spread better than countries that have a more dispersed urban form, like the US, Canada and some parts of Europe". Adlakha and Sallis[7] dismiss the role played by population density in spreading the disease on the basis of casual empiricism (a scatter plot that shows near-zero correlation between density and the spread of the disease in 36 cities). Li et al.[8] think that the evidence on possible links between population density and the spread of pandemics is inconclusive.

On 23 March 2020, the journalist Matt Yglesias, known to be sardonic and facetious, tweeted the following: "The moral of corona virus is that we should adopt the kind of low-density living patterns associated with Asian countries like South Korea, Japan, Taiwan, and Singapore that have successfully controlled its spread".[9] This tweet was meant to be sarcastic, perhaps directed at those blaming population density for making New York City the epicentre of the pandemic in the US. Barr and Tassier[9] argue that even though Asian cities are known for their "hyper-density", they have figured out ways to slow its spread without destroying the very essence of what makes cities so successful.

On the other hand, it is widely believed that population density is the main reason why New York City is the worst affected city in the US. This view is echoed by the author of an article entitled "Density is New York City's Big 'Enemy' in the Coronavirus Fight", which appeared in the *New York Times* on 23 March 2020. The author of the article refers to what he calls a "distinct obstacle in trying to stem new cases: its cheek-by-jowl density".[10] He insinuates that population density in New York City is the reason for the lack of success of the measures intended to slow the spread of the coronavirus, including the shutting down of non-essential businesses and urging its residents to stay home.

This divergence of views on the effect of density has led Barr and Tassier[9] to come up with the concept of the "density paradox", as they wonder about the role of population density in spreading the virus, whether big cities are more vulnerable than smaller ones, and if New York deserves "such harsh criticism". Based on a formula that defines the reproduction rate (the average number of people at the receiving end of transmission from an already infected person) and the assumption that no preventive measures (such as social distancing and self-isolation) are taken, they specify a regression equation whereby the number of cases is determined by several explanatory variables, including population density. By using county-level US data, they obtain results showing that "population density does matter but is not as large as the popular media would have you believe". More specifically, they find that (on average) an increase in a county's population density by 20% leads to a rise in the number of cases by about 11%–12% (conditional on having at least one case of the virus). They reach the conclusion that "more populous counties are likely to have fewer cases on a per capita basis than their sparser counterparts". They go on to argue that epicentres around the world are not located in the largest cities—for example, the epicentre in Italy is Milan (not Rome) whereas the Chinese epicentre is Wuhan (China's 10th largest city) rather than Shanghai, which is China's largest city.

This is such an important issue that warrants more scrutiny. The objective of this paper is to provide further evidence on the role played by population density in spreading the coronavirus, based on cross-sectional data covering 172 countries (obtained from several sources, including the European Centre for Disease Prevention and Control, the World Bank and the Center for Health Security). We will highlight the sensitivity of the results to the choice of the control variables and, unlike Barr and Tassier,[11] seek to identify the role of population density in determining both the numbers of cases and deaths per million of the population (as opposed to total cases and deaths). For this purpose, we use control variables that pertain to the ability of each country to deal with the pandemic as measured by the global health security (GHS) index, including measures of prevention, detection and reporting, rapid response, health system, compliance with international norms, and the risk environment.[12] The findings should have implications for health planning and management and also for urban planning.

1.1 | The issue under investigation and stylised facts

In 1854 a physician named John Snow was investigating the source of contamination leading to a widespread outbreak of cholera and killing more than 600 people in Soho, a London neighbourhood. He found that the culprit was a water pump located near the corner of Broadwick and Lexington. His endeavour to connect the dots between incidents of illness in the neighbourhood and the use of water from the pump was a feat of early empirical science that is heralded as a milestone in public health. Berg[13] describes the work of John Snow as providing a "textbook example of the link between disease and population density" and highlights the importance of this finding as a "link that's becoming even more important today as the world undergoes a dramatic process of urbanization".

The effect of population density works via the contact rate, which is a determinant of the reproduction rate (e.g.,[14]). The World Health Organisation (WHO) highlights the effect of population density by stating the following:[11]

For communities, inadequate shelter and overcrowding are major factors in the transmission of diseases with epidemic potential such as acute respiratory infections, meningitis, typhus, cholera, scabies, etc. Outbreaks of disease are more frequent and more severe when the population density is high.

The effect of population density on the spread of disease can be looked at from two different angels. On the one hand, densely populated regions lead to more face-to-face interaction among residents, which makes them potential hotspots for the rapid spread of pandemics. On the other hand, these regions may have better access to healthcare facilities and greater implementation of social distancing policies and practices. Therefore, the effect of population density on the spread of disease could be positive or negative. The empirical results also depend on which control variables are used.

The role played by population density in spreading pandemics and epidemics has been investigated repeatedly. For example, Tarwater and Martin[15] examine the effect of population density on the epidemic outbreak of measles or measles-like infectious diseases and find that a decline in a susceptible contact rate, from four to three, results in a "dramatic effect on the distribution of contacts over time, the magnitude of the outbreak, and, ultimately, the spread of disease".

Likewise, Maybery[1] attempts to identify the relationship between population density and the initial stages of the spread of disease. His results support the proposition that the number of new infections is strongly related to the distribution of susceptible contacts. Li et al.[8] examine possible links between population density and the propagation and magnitude of epidemics, arguing that it is inconclusive for three reasons: (i) a lack of focus on appropriate density intervals; (ii) for density to be a meaningful variable, the population must be distributed as uniformly as possible; and (iii) in propagation of an epidemic the initial proportion of susceptibles is an essential, yet usually unknown factor.

---

[1]WHO: What are the health risks related to overcrowding? https://www.who.int/water_sanitation_health/emergencies/qa/emergencies_qa9/en/.

A mathematical model is used by Sumdani et al.[2] to study MERS-CoV (Middle East Respiratory Syndrome-Coronavirus) and identify possible patterns of the spread. They split the population into two groups, with contact rates that are independent of and dependent on population density. By analysing the conditions under which the disease spreads, they observe how population density affects the transmission characteristics predicted by the epidemiological model.

Consistency between the basic epidemiological model and the effect of population density is demonstrated by Barr and Tassier[11] who start with the determinants of the reproduction rate, which include the contact rate (the average number of contacts people have with other people on a given day), the transmission rate (the likelihood that any one individual will pass the disease onto someone else), duration (the time it takes for the disease to work its way through the human body from infection to illness to recovery), and susceptibility (the fraction of people at a given time who are susceptible to the disease). Then they suggest that density affects the contact rate of individuals, which makes the reproduction rate larger, leading to more infections in dense regions. However, they emphasise the "missing role of time", in the sense that the spread of the disease at a particular point in time depends on when it started at that particular locality.

Other studies that deal with population density explicitly include Kadi and Khelfaoui,[3] who use cluster analysis to reveal strong correlation between population density and the number of infections, and Carozzi et al.[4] who find that density affects the timing of the outbreak, with denser locations more likely to have an early outbreak and that population density is not positively associated with time-adjusted cases and deaths. Tammes[16] finds that following the introduction of social distancing measures in England, the infection rate dropped more quickly in most densely populated regions. Sunn et al.[17] find that population density is not an important factor in Covid spreading under strict lockdown policies, arguing that the lockdown policies of China put a limit on the Covid-19 spreading speed.

As pointed out by Hsu[18] an important issue that seems to be overlooked in this debate is that "the population density of a city or county does not capture the finer points of how people actually gather within smaller spaces, such as those on college campuses or in individual residential buildings". Crowding occurs in densely as well as lightly populated cities. For example, crowding occurs when people gather for events, such as concerts and parties, and it can be caused by socioeconomic conditions that force many people to live in a small space or from cultural preferences for living in multi-generational households. Buses and other forms of mass transit can also get crowded, even in smaller urban areas. This is why Hsu[18] argues that "simple density has not adequately predicted the disease's course in the U.S., where the new coronavirus has spread well beyond urban areas to ravage rural communities and suburbs during the country's long summer".

The contrast in the results can be seen in the works of Moosa and Khatatbeh[19,20] and Hamidi et al.[21] Compared to demographic factors, Moosa and Khatatbeh[20] find that the infection rate depends on urban population rather than overall population density and that the mortality rate depends on the age structure of the population and population density but not on the percentage of urban population. Moosa and Khatatbeh[19] identify three robust variables that determine the case mortality rate: international tourist arrivals, population over 65 and population density. On the other hand, Hamidi et al.[21] find that county density is not significantly related to the infection rate, which they attribute to adherence to social distancing guidelines. However, they also find that counties with higher densities have significantly lower virus-related mortality rates than do counties with lower densities, which they attribute to "superior health care systems". They conclude that connectivity matters more than density in the spread of the pandemic and that large metropolitan areas with a higher number of counties tightly linked together through economic, social, and commuting relationships are the most vulnerable to the pandemic outbreaks.

Now a look at the stylised facts may be useful. In Figure 1 we observe population density, cases per million and deaths per million for the countries ranked highest in terms of the GHS index. These countries comprise a group of low-density ones (such as Australia, Canada and Norway) and high-density ones (such as Belgium, Netherlands and Korea). As expected, we cannot see one-to-one correspondence between rankings in terms of population density, cases and deaths, simply because population density is not the only factor determining cases and deaths.
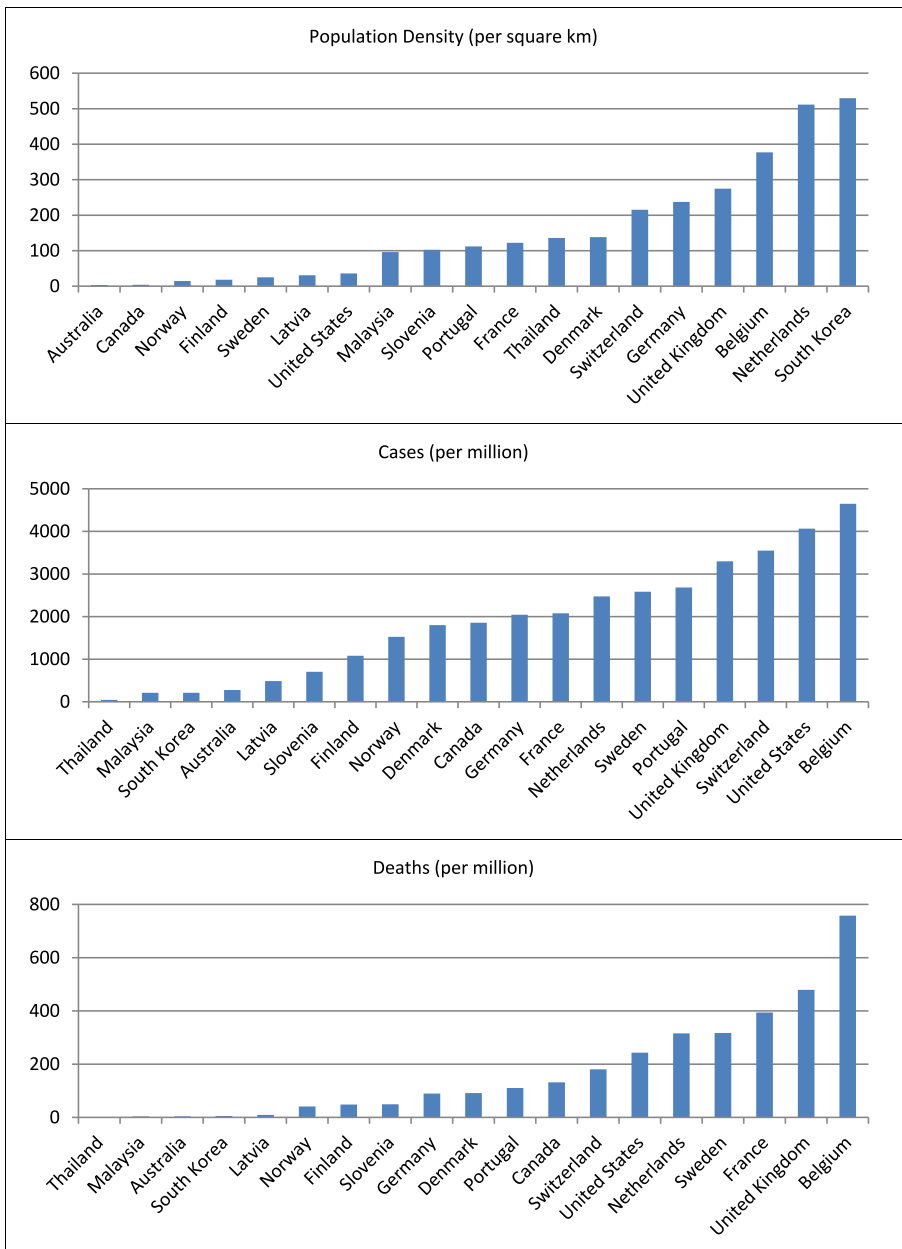
FIGURE 1 Ranking of countries with high GHS index. Note: These countries are selected on the basis of having high GHS indices and either high or low population density. Cases and deaths are totals as of 11 May 2020 [Colour figure can be viewed at wileyonlinelibrary.com]

The information embodied in Figure 1 is summarised in Table 1 and Table 2. Countries are classified into high, medium and low in terms of density, cases and deaths according to where they fall in the first, second or third 33 percentiles. Then the countries are placed into one of six cells ranging between high density-high cases/deaths and low density-low cases/deaths. If the number of cases per million is determined largely by density, most countries would be placed in the low-low and high-high cells in Table 1. We can see that six countries, out of a total of 19, fall in these cells. Yet, we find that two low-density countries (Sweden and the US) have high numbers of cases. This is because Sweden has not imposed measures of social distancing and perhaps because of mishandling and

TABLE 1 Classification of some countries by density and number of cases

	Population Density		
Cases per million	Low	Medium	High
High	Sweden	Portugal	Switzerland
	US		UK
			Belgium
Medium	Canada	France	Germany
	Norway	Denmark	Netherlands
Low	Australia	Malaysia	Korea
	Finland	Slovenia	
	Latvia	Thailand	

TABLE 2 Classification of some countries by density and number of deaths

	Population Density		
Deaths per million	Low	Medium	High
High	Sweden	France	UK
	US		Belgium
			Netherlands
Medium	Canada	Slovenia	Switzerland
		Portugal	Germany
		Denmark	
Low	Australia	Malaysia	Korea
	Norway	Thailand	
	Finland		
	Latvia		

complacency in the US. On the other hand, Korea has a high population density and low cases per million. This is due to the active and early implemented policy of test and trace. The pattern with respect to the number of deaths per million, as show in Table 2, is similar but not identical.

1.2 | Modelling issues

If population density leads to more cases and/or deaths, then we should expect, ceteris paribus, that countries with high densities record a higher incidence of cases and deaths. On the assumptions that the transmission rate, duration rate and susceptibility rate are likely to be constant across the US, Barr and Tassier[11] suggest the following regression equation to explain inter-county differences in the number of cases by population density:

$$ln(C) = \gamma_0 + \gamma_1 \, ln(D) + \boldsymbol{\delta}' \boldsymbol{X} + \varepsilon \tag{1}$$

where C is the total number of cases, D is population density and $\boldsymbol{X}$ is a set of control variables. However, they end up estimating a regression with the explanatory variables being population, land area, real GDP, days since first

case and airport passenger arrivals (Enplanements Mill+). It is not obvious why population and land area appear as separate explanatory variables when they should appear as a ratio defining density.

The problem with this specification is that the results are sensitive to the choice and definition of control variables. Young and Holsteen[22] argue that theory rarely says which variables should appear in the model, suggesting that "theory can be tested in many different ways and modest differences in methods may have large influence on the results". Likewise, Klees[23] believes that "all relevant variables that may affect the dependent variable can never be included". In fact, Young and Holsteen[22] are sceptical about the idea of using control variables because they (control variables) are "a common source of uncertainty and ambivalence". We hardly know what the "true model" is, in which case a misspecified model with 10 control variables is not superior to, or less biased than, a misspecified model with only five of them. Clarke[24] argues that extra controls can leverage correlations with other omitted variables, amplifying omitted variable bias. Likewise, Meng[25] refers to "an unknown number of relevant factors", some of which may be unknown to the economist, making it impossible to claim that all important variables are included in the model.

Gilbert[26] argues that published results are accepted because the reported regression equation has coefficients that are correctly signed and statistically significant. However, he suggests that these significant coefficients cannot be taken as evidence for or against the hypotheses under investigation, wondering about the other 999 regressions that have been assigned to the bin. This is what prompted Leamer[27] to suggest that econometricians confine themselves to publishing mappings from prior to posterior distributions rather than actually making statements about the economy.

A straight cross-sectional regression can be written as

$$Y = \alpha + \beta D + \sum_{i=1}^{n} \delta_i X_i + \varepsilon \tag{2}$$

where Y is the dependent variable (which can be cases per million or deaths per million), D is the variable of interest (population density) and X_i is one of n control variables, which are thought to affect the dependent variable. The problem here is that we can never be sure whether or not we have included all of the potential control variables, which is consequential as the results are sensitive to the set of control variables included in the model.

To circumvent this problem, Leamer[27,28] proposed extreme bounds analysis (EBA), which is a sensitivity analysis that enables the selection of the explanatory variables to be included in empirical models. A simple form of EBA involves the estimation of a series of regressions where the explanatory variables are the variable of interest (density) and a combination of h control variables, such that $1 < h < n$. Instead of estimating one equation containing all of the control variables, we estimate m equations, such that

$$m = n!/[h!(n-h)]! \tag{3}$$

For given values of n and h, the first and last equations are

$$Y = \alpha + \beta D + \delta_1 X_1 + \delta_2 X_2 + \ldots + \delta_h X_h + \varepsilon \tag{4}$$

And

$$Y = \alpha + \beta D + \delta_{n-h+1} X_{n-h+1} + \ldots + \delta_{n-1} X_{n-1} + \delta_n X_n + \varepsilon \tag{5}$$

In EBA, inference is based on the estimated coefficient, β, not from one equation, but rather from the whole set of estimates derived from the m estimated equations.

According to Leamer's EBA, D is a robust determinant of Y if the coefficient β does not change sign and significance. For example, when $n = 7$ and $h = 3$, β must remain statistically significant and of the same sign in 35 estimated regressions in order for the underlying variable of interest to be robust. This is a rather difficult test to

pass, particularly for high values of n, because if the coefficient changes sign and/or significance in one out of thousands of regressions, the variable is deemed fragile. The emphasis, however, shifts from significance (in one estimated equations) to robustness (in a large set of estimated equations).

To deal with the excessive stringency of Leamer's EBA, which also overlooks the distribution of β, Sala-i-Martin[29] has come up with an alternative EBA test. The alternative test involves the same procedure and number of regressions, but the criterion used to determine robustness is different. For this purpose, the entire distribution of β is analysed to determine the fraction of the cumulative distribution function (CDF) lying on each side of zero, CDF(0). If at least 95% of the CDF lies on either side of zero, the variable is considered robust—otherwise, it is fragile.

1.3 | Data and empirical results

The empirical results are based on a cross-sectional sample covering 172 countries.[2] Three dependent variables are used: cases per million, deaths per million and the death rate (deaths per case). While it is true that population density is more relevant to the determination of cases rather than deaths, it can be argued that it also affects the death rate, because hospitals are more likely to be overwhelmed in high than low density regions. Data on cases and deaths were obtained from the European Centre for Disease Prevention and Control.[3] The daily figures reported by the ECDPC are converted into cumulative totals as on 11 May 2020. Data on population density (measured in people per square kilometre) were obtained from the World Bank's development indicators data base.[4]

The control variables are taken to be measures of preparedness and the ability to deal with disease outbreak. These variables are the constituent components of the global health security (GHS) index, which is prepared jointly by the Nuclear Threat Initiative (NTI) and the Johns Hopkins Center for Health Security (JHU) (in co-operation with the Economist Intelligence Unit, EIU).[5] These variables are prevention (preventing the emergence or release of pathogens); detection and reporting (early detection and reporting for epidemics of potential international concern); rapid response (response to and mitigation of the spread of an epidemic), health system (sufficient and robust health system to treat the sick and protect health workers); compliance with international norms (commitment to improving national capacity, financing plans to address gaps, and adhering to global norms); and risk environment (the overall risk environment and country vulnerability to biological threats). These are certainly not the only control variables that can be used, but the idea is that they should reduce the effects exerted by population density on the dependent variable. This will help us answer the question whether, for example, the health system is good enough to counter the cases and deaths caused by population density. These variables (labelled $X_1, ...X_6$) are measured as indices taking values between zero and 100.

The EBA results are reported in Table 3 for both Leamer's EBA (looking at the extreme values only) and the Sala-i-Martin EBA (looking at the cumulative density function). The table displays the extreme values of the coefficient on the variable of interest (β_{min} and β_{max}), together with their t statistics in parentheses. The percentage of significant coefficients across regressions estimated for each combination of the control variables is also reported. We can see that β_{min} and β_{max} do not change sign or significance when the dependent variable is cases but not so when it is deaths. This means that population density is a robust determinant of cases, having a positive effect—that is, the virus spreads more quickly in more densely populated countries. Irrespective of the combination of control variables, the coefficient turns out to be significant (hence the 100%).

---

[2]The analysis is based on cross-sectional data where each country represents an observation in the sample. In this sense, the results do not pertain to individual countries, but rather to the sample as a whole. The sample consists of the latest available observation on each variable.
[3]https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide
[4]https://data.worldbank.org/indicator/EN.POP.DNST
[5]https://www.ghsindex.org/

TABLE 3 Robustness of population density

Sample	β_{min}	β_{max}	Significance (%)	CDF(0)
Cases/million	0.28	0.56	100.0	99.9
	(2.91)	(3.12)		
Deaths/million	−0.01	0.003	0.00	67.1
	(-1.41)	(0.22)		
Ratio (Deaths/Cases)	−0.000007	−0.000005	100.0	100.0
	(−3.50)	(−2.50)		

On the other hand, the coefficient on population density is not significant (even in one regression) when the dependent variable is the death rate, in which case it is definitely fragile. This means that population density does not affect the number of deaths, which makes sense because a larger number of infected people recover in a country with a high quality healthcare system. However, when the dependent variable is the ratio of deaths to infection cases, population density turns out to be robust, exerting a negative effect on the ratio. This means that with high population density, proportionately less of the infected people die. If cases rise with density but deaths are not affected by it, it follows that the ratio must decline as density rises.

The importance of density as a determinant of Covid-19 cases is demonstrated by the finding that it passes the difficult test of Leamer's EBA. No matter what combination of the control variables is used, the coefficient turns out to be significantly positive when the dependent variable is cases per million and negative when it is the ratio of deaths to cases. The results obtained from the Sala-i-Martin EBA are consistent, as the percentage of the cumulative density function lying on either side zero, CDF(0), is 100% in both cases.

2 | SUPPLEMENTARY RESULTS

To shed some light on the relative importance of the control variables as determinants of cases and deaths, we use variable addition tests and the Akaike information criterion for model selection. Now that we have found out that population density has a significant effect on the number of cases (and consequently on the ratio of deaths to cases), variable addition tests are intended to find out if any of the control variables adds anything in terms of explanatory power to an equation that has population density as the only explanatory variable. For this purpose, the following two regression equations are estimated:

$$Y = \alpha + \beta D + \varepsilon \tag{6}$$

$$Y = \alpha + \beta D + \delta_i X_i + \epsilon \tag{7}$$

where Y is any of the three dependent variables and X_i is any of the six control variables, such that $i = 1, ..., 6$. The null hypothesis is that X_i does not add anything to the explanatory power of the equation over and above what is provided by population density. For this purpose, three test statistics can be constructed from the residual sums of squares of equations (6) and (7): a Langrage multiplier (LM) statistic with a $\chi^2(1)$ distribution, a likelihood ratio statistic with a $\chi^2(1)$ distribution, and an $F(1,169)$ statistic. A significant test statistic implies that the added variable, X_i, does matter and that it contributes to the explanatory power of the equation (for details, see[30])

The Akaike information criterion (AIC) is used to choose between equation (6), in which the dependent variable is explained by population density only, and the following equation:

TABLE 4 Variable addition tests and model selection (cases per million)

Variable	Variable Addition Tests			Model Selection	
	LM	LR	F(1,169)	AIC[b]	Preferred Equation
X_1	6.20[a]	6.31[a]	6.32[a]	−0.46	8
X_2	1.53	1.53	1.52	2.05	6
X_3	3.26	3.29	3.27	0.87	6
X_4	9.79[a]	10.09[a]	10.21[a]	−2.22	8
X_5	0.49	0.49	0.49	2.59	6
X_6	34.96[a]	39.09[a]	43.12[a]	−18.1	8

[a]Significant at the 5% level, implying that the underlying variable is important and should be included in the model.
[b]A negative AIC implies that equation (8) (with the alternative variable) is preferable to equation (6) (with population density), and vice versa.

TABLE 5 Variable addition tests and model selection (deaths per million)

Variable	Variable Addition Tests			Model Selection	
	LM	LR	F(1,169)	AIC[b]	Preferred Equation
X_1	12.41[a]	12.88[a]	13.14[a]	−6.44	8
X_2	6.44[a]	6.56[a]	6.57[a]	−3.27	8
X_3	5.78[a]	5.88[a]	5.87[a]	−2.94	8
X_4	10.79[a]	11.14[a]	11.32[a]	−5.56	8
X_5	0.99	0.99	0.98	−0.44	8
X_6	27.18[a]	29.58[a]	31.72[a]	−14.62	8

[a]Significant at the 5% level, implying that the underlying variable is important and should be included in the model.
[b]A negative AIC implies that equation (8) (with the alternative variable) is preferable to equation (6) (with population density), and vice versa.

$$Y = \alpha + \delta_i X_i + \epsilon \qquad (8)$$

where the dependent variable is explained by one of the control variables. AIC is calculated as follows:

$$AIC = LL_1 - LL_2 - (k_1 - k_2) \qquad (9)$$

where LL_1 and LL_2 are respectively the maximised log-likelihood functions of equations (6) and (8), and k_1 and k_2 are the number of estimated parameters.[31,32] A positive value of AIC indicates that equation (6) is preferred to equation (8), and vice versa. When equation (6) is selected as the preferred model, this means that population density is more important for the determination of the dependent variable than the control variable X_i.

The results are presented in Tables 4, 5 and 6 for the three dependent variables. Table 4 reports the results when the dependent variable is cases per million. The variable addition test statistics show that three control variables add value over and above what is provided by population density: X_1 (prevention), X_4 (the healthcare system) and X_6 (the risk environment). These results are consistent with the preferred equation as selected by AIC, which tells us that X_1, X_4 and X_6 have better explanatory power than population density. In Table 5 we can see that any of the control variables has higher explanatory power for the number of deaths than population density. In Table 6, on the other hand, we can see that population density outperforms any of the control variables in explaining the ratio of death to cases.

TABLE 6 Variable addition tests and model selection (ratio of deaths to cases)

Variable	Variable Addition Tests			Model Selection	
	LM	LR	F(1,169)	AIC[a]	Preferred Equation
X_1	0.16	0.16	0.15	10.97	6
X_2	0.01	0.01	0.004	11.05	6
X_3	2.02	2.09	1.95	10.79	6
X_4	0.58	0.59	0.53	11.03	6
X_5	3.19	3.38	3.22	9.07	6
X_6	0.002	0.002	0.002	11.05	6

*means a significant test statistics, Significant at the 5% level, implying that the underlying variable is important and should be included in the model.

[a]A negative AIC implies that equation (8) (with the alternative variable) is preferable to equation (6) (with population density), and vice versa.

3 | CONCLUDING REMARKS

The "density paradox" refers to the observation that some highly populated cities and countries have recorded smaller numbers of Covid-19 cases than places that are sparsely populated. Controversy has arisen as to why New York City is the epicentre of Covid-19, which some observers attribute to high population density, whereas Seoul (which is more densely populated) has significantly fewer cases. On a country level, it can be readily seen that some sparsely populated countries have recorded more Covid-19 cases per capita than some more populated countries (e.g., Sweden vs. Slovenia). Yet, we can also observe that some sparsely populated countries have fewer cases than densely populated countries (e.g., Australia and Finland vs. Switzerland, the UK and Belgium).

This is not a paradox unless we make the implausible and unrealistic assumption that the only factor that determines Covid-19 cases per capita is population density. The outcome depends on an interplay of a large number of factors, including (but not limited to) population density. These factors include preparedness for epidemics and pandemics as well as the timing and nature of action taken to prevent the spread of the virus. For example, Sweden has the largest number of cases in Europe on a per capita basis because the government refuses to impose social distancing. Brazil, with a low density, is in a very bad shape because the president believes that Covid-19 is not more hazardous than the common cold (or flu at worst) and that the economy is more important.

In this paper we present empirical evidence on the role played by population density in spreading the coronavirus, based on cross-sectional data covering 172 countries. As control variables, we use measures of the ability of each country to deal with the pandemic as reflected in the global health security (GHS) index. Several empirical techniques are used, including extreme bounds analysis (both the Leamer and Sala-i-Martin versions), variable addition tests and the AIC as a model selection criterion. The results show that population density has a significantly positive effect on the number of cases but not the number of deaths—the latter is better explained by measures of preparedness. Population density also seems to have a negative effect on the ratio of deaths to cases. We can only conclude that the "density paradox" is not really a paradox.[6]

While views and findings differ on the role of population density in transmitting Covid-19, the effect of density on the spread of the disease is intuitive and consistent with the basic epidemiological model. This, however, does not mean that we should expect a one-to-one correspondence between density and the spread of the disease

---

[6]The words "puzzle" and "paradox" appear frequently in the economics literature to describe certain observations that are seemingly unexplainable (this does not necessarily mean that it does not have an explanation but that we do not know what the explanation is). However, it often turns out that these puzzles and paradoxes are not really puzzles or paradoxes (see, e.g.,[33]).

because other factors play a role as well. The results of this study show that, on a worldwide basis, density does matter. If this is the case, then what are the implications for health planning and management? One has to bear in mind that the objectives of healthcare in a pandemic are to minimise (i) transmissibility, morbidity and mortality; and (ii) the burden on the healthcare system. One implication is that if population density is conducive to the spread of the virus, then lockdowns and other forms of non-pharmaceutical intervention should not be uniform across regions of the same country. Rather the degree of stringency should depend on population density. However, public gatherings should be avoided even in lightly populated regions. Hsu[18] argues that not every crowding situation leads to widespread viral transmission, even though some have turned out to be superspreader events in a diverse array of settings in the US, including a suburban house party in Connecticut, a biotech conference at a Boston hotel, a Bible study session at a rural Arkansas Church, and overnight summer camps in Georgia and Missouri.

Population density has another implication because it has been found that middle-aged and older adults who live in densely populated neighbourhoods tend to have higher odds of being overweight, which makes them more vulnerable to Covid-19.[34] The effect goes through mode of travel and physical exercise, in the sense that those living in densely populated neighbourhoods typically use cars and spend less time on exercise. Yet another implication is that the pandemic could encourage some local governments and developers to promote suburban living. However, it is unlikely that there will be a mass exodus out of cities. Johnston[6] suggests that it is unlikely that Covid-19 alone will see a shift back to suburbanisation, even though some people may reconsider the desirability of some of the most high-density housing stock. He notes the view of an expert who says that "people won't necessarily want to move to the country after this, but there will be an acute awareness that they may need more space". The motivating factor is that people want to avoid riding in lifts and touching shared surfaces.

An important implication for health planning is highlighted by De Winter[35] who argues that what matters is not availability (e.g., the number of health centres per 10,000 people and the number of clinical laboratory per 100,000 people—the same goes for ICUs, doctors, nurses, etc.), but rather by accessibility, in the sense that people need health service within a certain distance from where they live. The geographical distribution of health personnel and facilities should not only depend on population density but also on accessibility.

The lesson that should be learned from these results is that density does matter, in which case it should not be dismissed as irrelevant on the basis of casual empiricism (e.g.,[7]) It is a factor that should be taken seriously by health planners and managers when they determine the distribution of health personnel and facilities. In the past, epidemics and pandemics provided incentives to do something constructive. e.g., London's sewage systems were developed in response to cholera outbreaks in the 19th century. This pandemic has provided the opportunity to do practical research that can aid decision making.

ETHICS STATEMENT

This study does not contain human or animal subjects.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available and can be obtained from several sources, including the European Centre for Disease Prevention and Control, the World Bank and the Center for Health Security.

ORCID

Ibrahim N. Khatatbeh https://orcid.org/0000-0001-7028-9661

REFERENCES

1. Maybery P. *The Effects of Population Density on the Spread of Disease*. Texas: Texas Medical Center Dissertations; 1999. https://digitalcommons.library.tmc.edu/dissertations/AAI9929469

2. Sumdani H, Frickle S, Le M, Tran M, Zaleta CK. *Effects of Population Density on the Spread of Disease*. Texas: University of Texas at Arlington; 2014. Technical Report 2014-05.

3. Kadi N, Khelfaoui M. Population density, a factor in the spread of COVID-19 in Algeria: statistic study. *Bulletin of the National Research Centre*. 2020;44(1):1–7. https://doi.org/10.1186/s42269-020-00393-x

4. Carozzi F, Provenzano S, Roth S. *Urban Density and Covid-19*. LONDON: CEP Discussion Papers; 2020. No 1711.

5. Wong DWS, Li Y. Spreading of COVID-19: Density matters. *PLOS ONE*. 2020;15(12):1–16. https://doi.org/10.1371/journal.pone.0242398

6. Johnston P. *Planning in an Era of COVID-19 and High Density Living*; 2020. The Fifth Estate, 14 April.

7. Adlakha D, Sallis JF. Activity-friendly neighbourhoods can benefit non-communicable and infectious diseases. *Cities & Health*. 2020;1–5. https://doi.org/10.1080/23748834.2020.1783479

8. Li R, Richmond P, Roehner BM. Effect of population density on epidemics. *Phys A Stat Mech its Appl*. 2018;510:713-724.

9. Barr JM, Tassier T. *Escape from New York? Density and the Coronavirus Trajectory*. 20, 2020. https://buildingtheskyline.org/covid19-and-density/

10. Rosenthal B. *Density Is New York City's Big 'Enemy' in the Coronavirus Fight*. New York Times. 2020:23.

11. Barr JM, Tassier T. *Modeling the Impact of Density on the Spread of the Corona Virus*, Working Paper. 2020. https://buildingtheskyline.org/wp-content/uploads/2019/10/Theory-and-Empirical-Model-6April20.pdf

12. Center for Health Security. *Global Health Security Index*. Baltimore: Johns Hopkins Bloomberg School of Public Health; 2019.

13. Berg N. *As Cities Grow More Crammed and Connected, How will we Discourage the Spread of Disease?* New York, NY. 2016. https://ensia.com/features/as-cities-grow-diseases-spread-faster-and-urban-design-key/

14. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Math Biosci*. 2013;244:125-134.

15. Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6:29-36.

16. Tammes P. Social distancing, population density, and spread of COVID-19 in England: a longitudinal study. *BJGP Open*. 2020;4(3).

17. Sun Z, Zhang H, Yang Y, Wan H, Wang Y. Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China. *Sci Total Environ*. 2020;746.141347

18. Hsu J. *Population Density Does Not Doom Cities to Pandemic Dangers*. New York, NY: Scientific American; 2020.

19. Moosa IA, Khatatbeh IN. International tourist arrivals as a determinant of the severity of Covid-19: international cross-sectional evidence. *Journal of Policy Research in Tourism, Leisure and Events*. 2020;1–16. https://doi.org/10.1080/19407963.2020.1859519

20. Moosa IA, Khatatbeh IN. Robust and fragile determinants of the infection and case fatality rates of Covid-19: international cross-sectional evidence. *Appl Econ*. 2021;53:1225-1234.

21. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? *J Am Plan Assoc*. 2020;86:495-509.

22. Young C, Holsteen K. Model uncertainty and robustness. *Socio Methods Res*. 2017;46:3-40.

23. Klees SJ. Inferences from regression analysis: are they valid? *Real-World Econ Rev*. 2016;74:85-97.

24. Clarke KA. The phantom menace: omitted variable bias in econometric research. *Confl Manag Peace Sci*. 2005;22:341-352.

25. Meng S. *Patentism Replacing Capitalism: A Prediction from Logical Economics*. New York: Palgrave-Macmillan; 2019.

26. Gilbert CL. Professor Hendry's econometric methodology. *Oxf Bull Econ Statistics*. 1986;48:283-307.

27. Leamer E. Let's take the con out of econometrics. *Am Econ Rev*. 1983;73:31-43.

28. Leamer E. Sensitivity analyses would help. *Am Econ Rev*. 1985;75:308-313.

29. Sala-i-Martin X. I just ran two million regressions. *Am Econ Rev*. 1997;87:178-183.

30. Pesaran B, Pesaran MH. *Time Series Econometrics: Using Microfit*. 5.0. Oxford: Oxford University Press; 2009.

31. Akaike H. Information theory and the extension of the maximum likelihood principle. In: Petriv B, Caski BF, eds. *Proceeding of the Second International Symposium on Information Theory*; 1973.

32. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19:716-723.

33. Moosa IA. *Controversies in Economics and Finance: Puzzles and Myths*. Cheltenham: Edward Elgar; 2020.

34. Wang R, Feng Z, Xue D, et al. Exploring the links between population density, lifestyle, and being overweight: secondary data analyses of middle-aged and older Chinese adults. *Health Qual Life Outcome*; 2019;17(1):1–10.

35. De Winter ER. Are we ignoring population density in health planning? The issues of availability and accessibility. *Health Pol Plann*. 1992;7:191-193.

AUTHOR BIOGRAPHIES

Imad N. Moosa is a retired professor of finance at RMIT University, Australia. He has published 25 books and over 250 papers in various areas of economics and finance. His latest book is "The Economics of Covid-19: Implications for Economic Thought and Public Policy" (in press).

Ibrahim N. Khatatbeh is an assistant professor of finance and risk management at the Hashemite University, Jordan. He Obtained his PhD in 2020 from RMIT University, Australia. He is interested in multidisciplinary research and his work has been published in several fields including finance, insurance, and health policy.