

Yes, Lockdown: What Is Hidden in COVID-19 Cross-Country Data

Pavel Jelnov*

April 21, 2020

Abstract

Measures to contain COVID-19 rely on a body of local estimates of the infection's contagiousness and fatality. I show that one can use cross-country data to estimate the upper bound of morbidity and lower bound of mortality. I rely on two strong cross-country correlations, i.e., between log of tests and log of recorded cases and between log of recorded cases and log of deaths. Using extrapolation, I find that the true infection rate is not higher than a few percent, and that cross-country mean case fatality is at least 2.5%. I conclude that the severe non-pharmacological interventions are justified.

JEL code: I18

Keywords: COVID-19 morbidity, COVID-19 mortality

The measures taken worldwide to contain the COVID-19 pandemic evolve as an unprecedented in the modern history economic drama. Yet the reported morbidity data is suspected

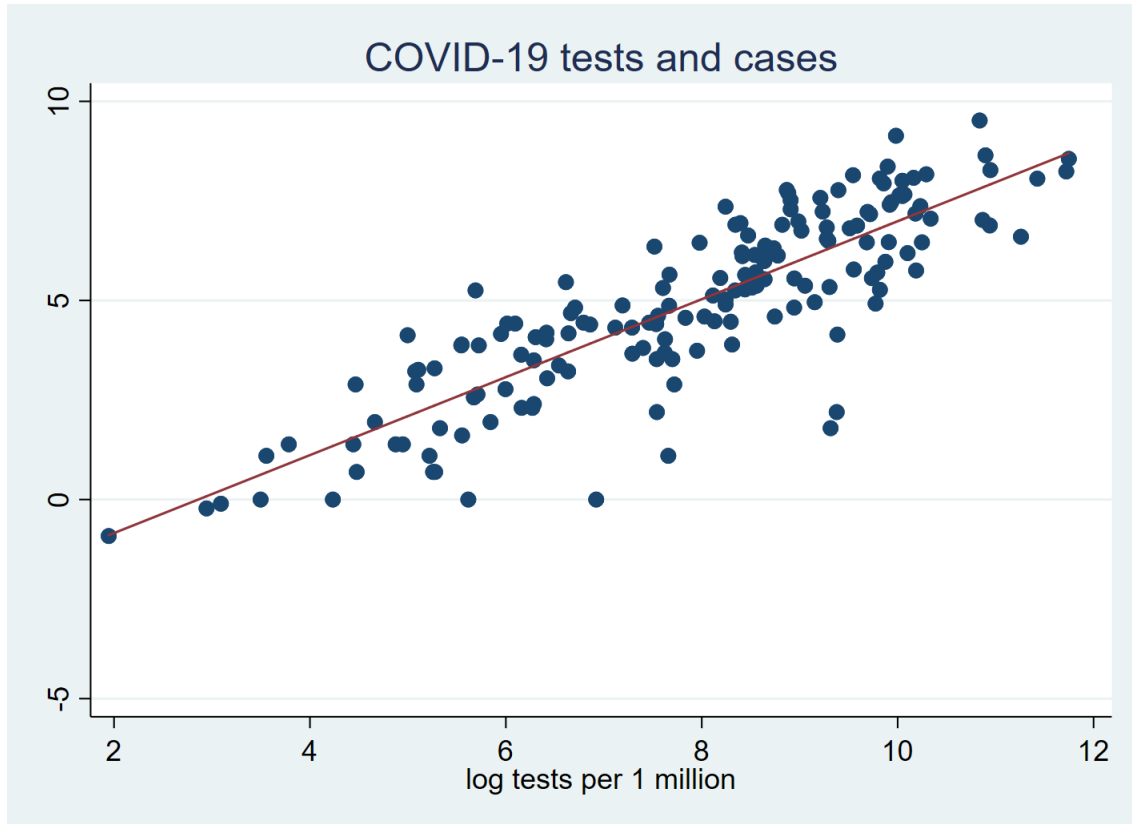
*Institut für Arbeitsökonomik, Leibniz Universität Hannover, and IZA, e-mail: jelnov@aoek.uni-hannover.de.

to be biased by country-specific testing policy, fraud, and mysterious properties of the virus. The unknown true levels of contagiousness and mortality have crucial implications on the effectiveness of non-pharmacological interventions. The attempts to use local data to estimate contagiousness and mortality continue (Qiu et al, 2020). The big question is whether the number of deaths from COVID-19 is relatively modest because the virus is not very contagious or because it is not very dangerous. Those who claim that COVID-19 is very contagious but not very dangerous overlook the reported morbidity rates and doubt the health benefits of the economically devastating measures. For instance, in the early stages of the crisis, the United Kingdom intended not to shut the economy down and to rely on development of “herd immunity”.

However, the cross-country data that links COVID-19 testing to recorded morbidity and recorded morbidity to mortality is far from being erratic. By contrast, it is well structured. In the simple analysis below, I use this fact to show that one month after COVID-19 was declared pandemic by World Health Organization (WHO, 2020), the virus is not spread among more than few percent of the world population. However, the virus is dangerous: the case fatality is, up to country-specific error term, at least 2.5%. To obtain these estimates, I predict the infection rate and the case fatality in a hypothetical country that tests all of its population. By the observed positive correlation between testing and infection rate, this country is also expected to have the highest infection rate.

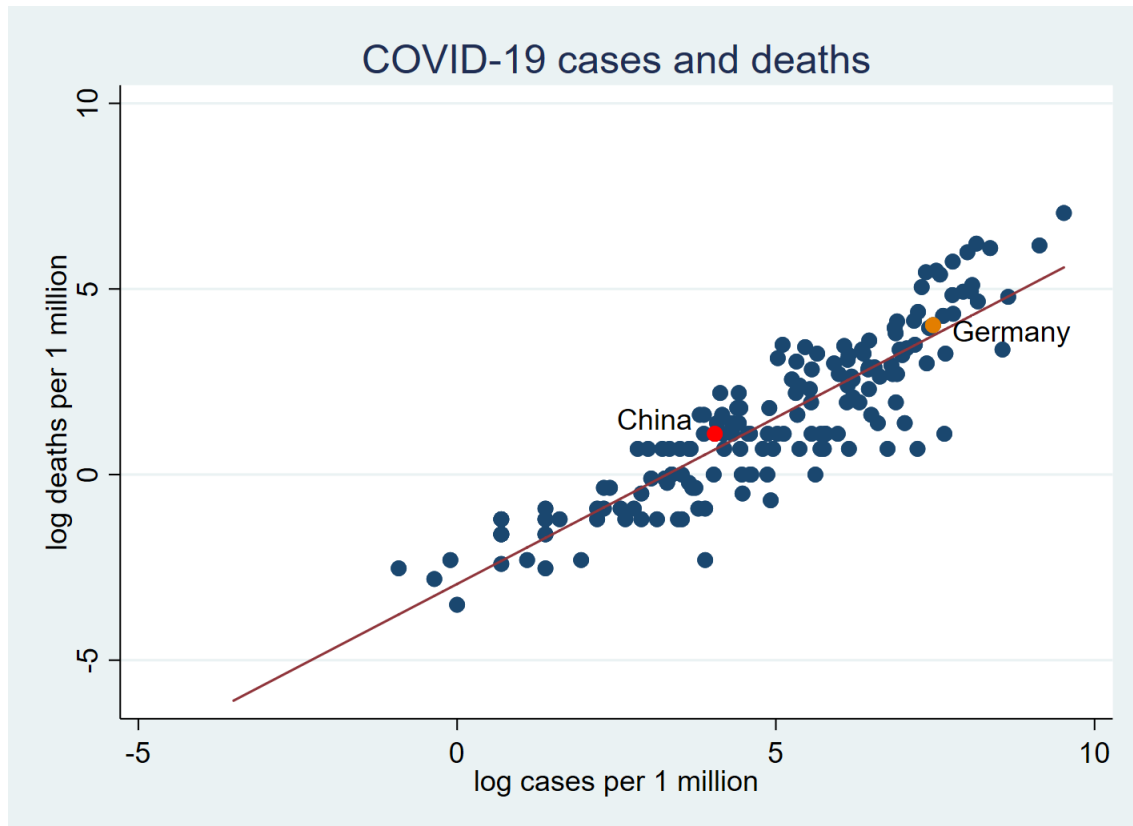
In detail, I observe two very strong cross-country correlations. First, the correlation between log of tests and log of reported cases (per capita) is 0.85 (Figure 1). Second, the correlation between log of reported cases and log of reported deaths (per capita) is 0.89 (Figure 2). Thus, for some structural reasons, the logs of test, recorded infection, and death

Figure 1: COVID-19 tests and reported cases



Note: The Figure presents log of COVID-19 tests per one million of population vs log of reported cases per one million of population in 163 countries. Data was released from <https://www.worldometers.info/coronavirus/> on April 20, 2020.

Figure 2: COVID-19 reported cases and deaths



Note: The Figure presents log of COVID-19 reported cases per one million of population vs log of reported deaths per one million of population in 163 countries. Data was released from <https://www.worldometers.info/coronavirus/> on April 12.

rates follow a worldwide mechanism despite all the differences between countries. A simple model that links these variables is

$$\log(\text{cases}_i) = \alpha_0 + \alpha_1 \cdot \log(\text{tests}_i) + \varepsilon_i, \quad (1)$$

$$\log(\text{deaths}_i) = \beta_0 + \beta_1 \cdot \log(\text{cases}_i) + \nu_i, \quad (2)$$

where tests, recorded cases, and deaths are per capita, and ε_i and ν_i are country-specific error terms. This model is theoretically incomplete, because it assumes a single causal link in each equation. In particular, it ignores reverse causality of infection rate on testing intensity, and ignores the effect of an omitted variable “time since patient zero” (or patient 30, see Caspi et al., 2020) on all three variables. However, because of the strong correlations, the model is a good proxy for the purpose of prediction. Summary statistics of the three variables are shown in Table 1 (for convenience, the summary statistics and the axes in Figures 1 and 2 are re-scaled to per million of population).

Table 2 presents the results of ordinary least squares estimation of Equations (1) and (2). The estimated α_1 equals 0.979. Moreover, the hypotheses that α_1 equals one and that ε_i is homoscedastic cannot be rejected. Thus, we observe a constant proportion of positive out of all tests, up to country-specific multiplier with median one (under log-normality of ε_i). The estimated β_1 equals 0.896. Therefore, mortality is concave with regard to the recorded infection rate.

I estimate the expected maximal cross-country infection rate by extrapolation to full testing. Because α_1 is positive, the highest expected infection rate is associated with the highest possible testing intensity. Thus, by the hidden economic structure observed in Figures

Table 1: Summary statistics

| Variable | Mean | Std. Dev. | N |
|---------------------------|-------|-----------|-----|
| log(tests per 1 million) | 7.886 | 1.984 | 166 |
| log(cases per 1 million) | 4.605 | 2.397 | 210 |
| log(deaths per 1 million) | 7.886 | 1.984 | 166 |

Notes: The table presents summary statistics of the used variables. Data was released from <https://www.worldometers.info/coronavirus/> on April 20, 2020.

Table 2: OLS regressions

| VARIABLES | (1) ln(cases pc) | (2) ln(deaths pc) |
|--------------|----------------------|----------------------|
| ln(tests pc) | 0.979*** (0.0467) | |
| ln(cases pc) | | 0.896*** (0.0354) |
| Constant | -3.091*** (0.292) | -4.386*** (0.325) |
| Observations | 166 | 165 |
| R-squared | 0.728 | 0.797 |
| Root MSE | 1.1859 | 0.9986 |

Notes: The table presents OLS regression results of log cases per capita on log tests per capita and of log of deaths per capita on log of cases per capita. Data was released from <https://www.worldometers.info/coronavirus/> on April 20, 2020. Standard errors are given in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

1 and 2, the highest expected infection rate would be in a hypothetical country that tests all of its population. By coincidence, also the true number of cases can be revealed when all of the population is tested. Thus, to estimate the maximal expected true infection rate, one can set $tests = 1$. The point estimate of maximal $\ln(infection\ rate)$ is $\alpha_0 = -3.091$ (95% CI [-3.667, -2.515]). The expected infection rate in a full-testing country would be $E(infection\ rate) = Exp(\alpha_0 + \frac{\sigma_\varepsilon^2}{2}) = 0.092$ (95% CI [0.052, 0.163]). However, all countries are far below full testing, and, up to ε_i , far below the infection rate expected at the point where they decide on full testing. Therefore, the upper bound of the true infection rate is somewhere between 5% and 15%, probably closer to former than to latter.

Furthermore, because of concavity of Equation (2), the maximal recorded infection rate is associated with the lowest mortality. Thus, the lower bound of mortality is, up to ν_i , the predicted mortality for the maximal infection rate.

From column (2) of Table 2, the expected death rate when infection rate is 5% is $E(death\ rate) = Exp(\beta_0 + \beta_1 \ln(0.05) + \frac{\sigma_\nu^2}{2}) = 0.0014$ per capita, and COVID-19 case fatality is $\frac{0.0014}{0.05} = 0.028$. When infection rate is 15%, the predicted death rate is 0.0037 deaths per capita, and the case fatality is 0.025. Thus, because the upper bound of the expected true infection rate is somewhere between 5% and 15%, the case fatality lower bound, up to country specific-error term, is about 2.5%. This case fatality is not far away from estimates based on local data (Qiu et al, 2020). As a reference, the current¹ recorded case fatality (the share of deaths out of all “closed” recorded cases) in China, the first country to face COVID-19, is 6%. In Germany, a country with the highest number of “closed” recorded cases, it is 5%. Because around half of the COVID-19 cases are asymptomatic², and because China and

¹Obtained on April 20, 2020, from <https://www.worldometers.info/coronavirus/>.

²<https://www.dailymail.co.uk/news/article-8210401/Iceland-finds-half-population-asymptomatic->

Germany are located almost exactly on the regression line (see Figure 2), I conclude that my estimate makes sense: it is half of the recorded death rate in Germany and China. Moreover, the real case fatality is somewhat higher than any estimate based on the current number of deaths, because unfortunately some of those currently infected but still alive will die in the future.

To conclude, simple cross-country correlations reveal a systematic relationship between testing and recorded morbidity and between recorded morbidity and mortality. The correlations are strong enough to be used for prediction. Extrapolation generates upper bounds of morbidity and lower bounds of mortality rates that show that the virus is not extremely contagious but is dangerous. These results advocate the strict non-pharmacological interventions.

References

Caspi, Gil and Shalit, Uri and Kristensen, Soren Lund and Aronson, Doron and Caspi, Lilac and Rossenberg, Oran and Shina, Avi and Caspi, Oren, “Climate effect on COVID-19 spread rate: an online surveillance tool,” 2020.

Yun Qiu, Chen Yun, and Shi Xi, “Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China,” *Journal of Population Economics*, Issue 4, 2020.

World Health Organization, “WHO Director-General’s opening remarks at the media briefing on COVID-19 – 11 March 2020.” 11 March 2020.

infected-Covid-19.html