Cross-Country Data

Pavel Jelnov*

April 18, 2020

Abstract

Measures to contain COVID-19 rely on a body of local estimates of the infection’s contagiousness and fatality. I propose a simple method to estimate the cross-country upper bound of morbidity and lower bound of mortality using 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 in no country is above 8.5%, and that mean case fatality is at least 0.9%. I conclude that the severe policy measures are justified.

JEL code: I18

Keywords: COVID-19 morbidity, COVID-19 mortality

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

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 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 lockdown and social distance policy. The attempts to use local data to estimate contagiousness and mortality continue (see the analysis and the most up-to-date review of COVID-19 contagiousness estimation in 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 suggest that large parts of the population have been already infected, but have not developed symptoms and were not tested. Therefore, they overlook the reported morbidity rates and doubt the health benefits of the economically devastating lockdown. For instance, in the first 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 8.5% of the population in any country. This upper bound estimate conservatively takes into account the 95% confidence interval; the point estimate of the maximal cross-country infection rate is only 4.7%. However, the virus is dangerous: the case fatality is, up to country-specific error term, at least 0.9%.
To obtain the upper bound of infection rate, I predict the infection rate in a hypothetical country that tests all of its population. By the observed positive correlation between testing and infection rate, this country should also have the highest infection rate. Then I rely on concavity of the death rate with respect to the reported infection rate to predict mortality when the true infection rate is the cross-country maximum upper bound. This is the lower bound mortality estimate.

**Morbidity**

I observe two very strong cross-country correlations: between log of tests and log of recorded cases and between log of recorded cases and log of deaths (see the summary statistics of these variables in Table 1). First, the correlation between log of tests and log of reported cases (per capita) is 0.85. This correlation is shown in Figure 1. Column (1) of Table 2 presents the results of an OLS regression of the model

\[
log(\text{cases}_i) = \beta_0 + \beta_1 \cdot log(\text{tests}_i) + \epsilon_i;
\]

where cases and tests are per capita, and \(\epsilon_i\) is a country-specific error term. This model is theoretically incomplete, because it assumes a single causal link and, in particular, ignores reverse causality of infection rate on testing intensity. However, because of the strong correlation, it is a good proxy for the purpose of prediction. The estimated slope \(\beta_1\) is 0.975 (interestingly, the hypothesis that it equals one cannot be rejected).

I estimate the maximal cross-country infection rate by extrapolation to full testing. Because the slope is positive, the highest possible infection rate is associated with the highest
Table 1: Summary statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(tests per 1 million)</td>
<td>7.767</td>
<td>1.963</td>
<td>163</td>
</tr>
<tr>
<td>log(cases per 1 million)</td>
<td>4.476</td>
<td>2.432</td>
<td>210</td>
</tr>
<tr>
<td>log(deaths per 1 million)</td>
<td>1.277</td>
<td>2.263</td>
<td>163</td>
</tr>
</tbody>
</table>

Notes: The table presents summary statistics of the used variables. The data is taken from https://www.worldometers.info/coronavirus/ on April 17, 2020.

possible testing. By coincidence, also the true number of cases can be revealed when all of the population is tested. Thus, to estimate the maximal true infection rate, one can set $tests = 1$. The predicted maximal infection rate is $Exp(-2.81) = 0.047$. The 95% confidence interval is [0.025,0.085]. Thus, the point estimate of the cross-country maximal infection rate as 4.7% and the conservative upper bound of the infection rate 8.5%.

Mortality

Second, the correlation between log of reported cases and log of reported deaths (per capita) is 0.89. This correlation is illustrated in Figure 2. Once again, the cross-country data is systematic despite cultural and institutional differences between countries. The regression model is

$$
\log(deaths_i) = \alpha_0 + \alpha_1 \cdot \log(cases_i) + \nu_i, \quad (2)
$$

where deaths and cases are per capita, and $\nu_i$ is a country-specific error term. The OLS results\textsuperscript{1} are reported in column (2) of Table 2. The estimated slope $\alpha_1$ is 0.894. Thus, the

\textsuperscript{1} I do not estimate simultaneously Equations (1) and (2) in order not to lose observations: some countries do not report the number of tests, while some other countries report zero deaths.
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 taken from https://www.worldometers.info/coronavirus/ on April 17, 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 taken from https://www.worldometers.info/coronavirus/ on April 12.
Table 2: OLS regressions

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(cases per 1 million)</td>
<td>0.975***</td>
<td>0.894***</td>
</tr>
<tr>
<td></td>
<td>(0.0485)</td>
<td>(0.0353)</td>
</tr>
<tr>
<td>log(cases per 1 million)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2.723***</td>
<td>-2.984***</td>
</tr>
<tr>
<td></td>
<td>(0.388)</td>
<td>(0.186)</td>
</tr>
<tr>
<td>Observations</td>
<td>163</td>
<td>163</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.715</td>
<td>0.800</td>
</tr>
</tbody>
</table>

Notes: The table presents OLS regression results of log cases per one million of population on log tests per one million of population and of log of deaths per one million of population on log of cases per one million of population. Data was released from https://www.worldometers.info/coronavirus/ on April 17, 2020. Standard errors are given in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
death rate is concave with respect to the number of treated cases. This concavity means that the share of lethal cases decreases in the number of treated ones. In other words, the more COVID-19 cases a country treats, the lower the share of serious cases out of all treated is.

Most importantly, the strong correlation in Figure 2 shows that one can estimate the lower bound of mortality using extrapolation. The real mortality is \( \frac{\text{death rate}}{\text{infection rate}} \). The maximal true infection rate is in the range calculated in the previous section. Because of concavity of Equation (2), the maximal number of treated cases is also associated with the lowest mortality. Thus, the lower bound of mortality is, up to the confidence interval of the prediction and country-specific error term, the predicted mortality for the maximal infection rate.

The predicted mortality when the infection rate is 4.7%, the maximal infection rate point estimate, is 1.6%. However, to obtain the lower bound estimate for mortality, I consider the upper bound of the 95% confidence interval for the maximal infection rate, which is 8.5%. The corresponding 95% confidence interval of mortality is [0.009, 0.025]. Thus, COVID-19 cross-country mean case fatality is probably at least 0.9%. Moreover, the real mortality 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. As a reference, the current² reported mortality (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%. In Iceland, a country that performed the highest number of tests per capita (except of Faeroe Islands), it is 1%.

²Obtained on April 17, 2020, from https://www.worldometers.info/coronavirus/.
Conclusions

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 lockdown policy.

References
