March 23, 2020

Coronavirus Data Gaps and the Policy Response to the Novel Coronavirus

James H. Stock
Department of Economics, Harvard University

This note lays out the basic SIR epidemiological model of contagion, with a target audience of economists who want a framework for understanding the effects of social distancing and containment policies on the evolution of contagion and interactions with the economy. The model is calibrated to the most recent data, however it simple nature abstracts from many important considerations and its output is not intended to supersede estimates from more sophisticated epidemiological models.

This note makes four main points:

- The effect of social distancing and business shutdowns on epidemic dynamics enters the model through a single parameter, the case transmission rate β. For a specified case transmission rate, the policy design question is how to achieve that case transmission rate while minimizing economic cost. A second economic question is, what is the optimal case path for β, trading off the economic cost of that path against the costs in deaths.
- The parameters of the model are not well estimated in the literature on the coronavirus because of the lack of available data. Data on prevalence, for example, is obtained from positive rates of testing for the coronavirus, however so far tests have been rationed and almost entirely have been administered to a selected population, those at risk and showing symptoms. Thus, the fraction of tests that are positive do not estimate the population rate of infection.
- Using Bayes Law, it is possible to re-express the model in terms of β and the asymptomatic rate, which is the fraction of the infected who show sufficiently mild, or no, symptoms so that they are not tested under current testing guidelines. The virtue of re-expressing the model this way is that it makes use of the positive testing rate, on which there is good data. The COVID-19 asymptomatic rate is unidentified in our model and recent estimates in the epidemiological literature range from 0.18 to 0.86. However, the asymptomatic rate could be estimated accurately and quickly by testing a random sample of the overall population.
- The policy response and its economic consequences hinge critically on the asymptomatic rate. As we illustrate using two policy paths for β, without better knowledge of this knowable parameter, policymakers could make needlessly conservative decisions which would have vast economic costs.

A simple calibrated epidemiological model

Under the simplifying assumptions that the population mixes homogeneously, that the asymptomatic are as infectious as the symptomatic (possibly not true, see Li et. al. (2020), that the population is equally susceptible to contagion, and that those who have been previously infected are no longer susceptible. the infection rate follows the so-called SIR model (see for example Allen (2017)). The simple SIR model used here abstracts from mortality. The discrete-time version of the SIR model at the weekly time scale is:

$$\Delta S_t = -\beta I_{t-1} \frac{S_{t-1}}{N} \tag{1}$$

$$\Delta R_t = \gamma I_{t-1}, \tag{2}$$

$$\Delta I_t = \beta I_{t-1} \frac{S_{t-1}}{N} - \gamma I_{t-1} \tag{3}$$

where S_t is the number of susceptible, I_t is the number of infected, R_t is the number of recovered, and N is the (constant) total population. Assuming that everyone in the population is initially susceptible, then $N = S_t + I_t + R_t$. The coefficient β is the transmission rate and γ is the recovery rate.

Equation (1) says that the number of newly infected is the number of prior infected times the transmission rate times the fraction of the population that is susceptible; the number of susceptibles is reduced one-for-one by the number of newly infected. Equation (2) says that the number of recovered increases by the number recovered in the current period. Equation (3) says that the current number of infections increases by the number of new infections, minus the number of recoveries.

A common summary of disease transmission is the basic reproduction number, R_0, which is $R_0 = \beta/\gamma$. R_0 is the total number of cases produced by contagion from a single case, when the entire population is susceptible and β and γ are at their no-policy values. Policies that decrease β and/or increase γ serve to reduce R_0.

In this model, policy operates by manipulating the values of the parameters. The baseline values can be considered no-policy values. Self-quarantine, social distancing, business closures act to reduce the transmission rate β. Health interventions, such as medical treatment or drugs (should they become available) could serve to increase the recovery rate γ.

Parameter values and observable implications

The model has two unknown parameters, β and γ. For the coronavirus, surprisingly little data exists to estimate β and γ because testing has been limited and the testing that has been done has largely been targeted to the sick, especially the sick who are either the most vulnerable or who might benefit the most from hospitalization. That is, testing has largely been of the symptomatic. Such testing guidelines miss cases that are variously referred to as asymptomatic or simply undetected, which I take here to be synonymous. A case can be undetected because the individual has no symptoms, because she has sufficiently mild symptoms (cold or allergy symptoms) that she did not think to report the case, or because she reported her symptoms to a medical professional but did not meet strict guidelines for receiving the test.

Estimation of the model in a conventional sense, that is, fitting (1) - (3) using time series data, is not possible because there are no data on I_t and R_t with which to fit the model. Obtaining estimates of I_t would require ongoing random testing of the population, which has not happened. Similarly, estimates of R_t could be deduced from I_t and, given γ, or alternatively could be obtained by ongoing random sampling of

tests for serum antibodies in response to the coronavirus, however such tests do not currently exist although they are under development.

The absence of random testing of the population poses an additional problem. Although the recovery rate γ for the seriously ill can be estimated from data on those whose disease progression has been tracked, it is not estimated among the asymptomatic. The recovery rate plausibly differs among the symptomatic and asymptomatic, complicating direct estimation of γ from medical case data.

Although there are no data on I_t and R_t, there are widely available data on the results of testing (e.g., Roser, Ritchie, and Ortiz-Ospina (2020)). Because testing in the United States has largely focused on the symptomatic (putting aside small nonrepresentative asymptomatic groups like NBA players), it is plausible that the positive testing rate estimates the rate of infection among the symptomatic. Using Bayes law, the model can be augmented to take advantage of time series data on the positive testing rate.

The positive testing rate can be used to calibrate the SIR model in the following way. Dividing both sides of (1) - (3) by N expresses all quantities as population rates or, at the individual level, probabilities. Under the simplifying assumption that only the symptomatic are tested, we can use Bayes law to express the positive testing rate in terms of the symptomatic rate (the fraction of infections that are symptomatic):

$$\Pr\left[I_t \middle| Symptomatic_t\right] = \frac{\Pr\left[Symptomatic_t \middle| I_t\right]\Pr\left[I_t\right]}{\Pr\left[Symptomatic_t\right]} = \left(1-\pi_a\right)\frac{\Pr\left[I_t\right]}{\Pr\left[Symptomatic_t\right]} \tag{4}$$

where I_t and S_t refer to the infected and susceptible as above and where $\pi_a = \Pr\left[Asymptomatic_t \middle| I_t\right]$ $= 1 - \Pr\left[Symptomatic_t \middle| I_t\right]$ is the asymptomatic rate (the undetected infection rate).

The fraction of the population that is symptomatic (the denominator in (4)) is,

$$\begin{aligned}\Pr\left[Symptomatic_t\right] &= \Pr\left[Symptomatic_t \middle| I_t\right]\Pr\left[I_t\right] + \Pr\left[Symptomatic_t \middle| S_t\right]\Pr\left[S_t\right] \\ &= \left(1-\pi_a\right)\Pr\left[I_t\right] + s_0 \Pr\left[S_t\right],\end{aligned} \tag{5}$$

where s_0 is the baseline rate of symptoms among the susceptible.

Assuming that testing has been random among the symptomatic, the fraction of tests that are positive estimates $\Pr\left[I_t \middle| Symptomatic_t\right]$. The expanded system (1) - (5) has five equations and four parameters: β, γ, π_a, and s_0.

I do not explore estimation of the model here using time series data on the positive testing rate. Instead, I illustrate its use and the importance of the key parameter π_a, the asymptomatic rate, for policy, in a calibrated simulation.

Model calibration and simulation.

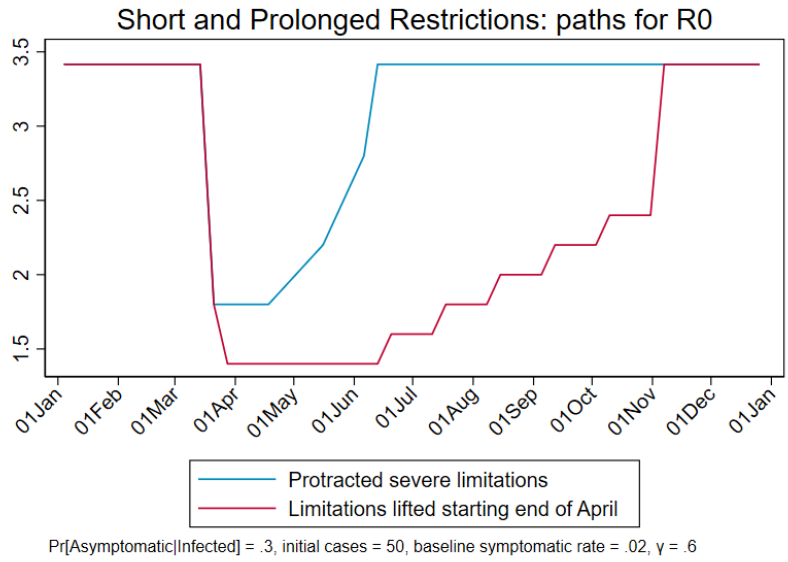
I now turn to an illustrative illustration of the model and policy interventions. For γ, I assume that the half-life of an infection is one week ($\gamma = 0.5$). I set $s_0 = 0.02$. For the week of March 21, 2020, the positive testing rate in the United States is approximately 10% (Roser, Ritchie, and Ortiz-Ospina (2020)). as initial conditions, I assume that there were 50 (unknown) cases in the US in the week ending January 4, 2020. Even with these calibrations, the model (1) - (5) is underidentified by one parameter. I therefore fix the asymptomatic rate π_a at a predetermined value and allow the model to solve for β such that the positive testing rate is 10% for the week ending March 21, 2020.

The limited available evidence on the asymptomatic rate has been reviewed in Qui (March 20, 2020) An early estimate of the asymptomatic rate suggested it could be as low as 18%, however that study used data from the *Diamond Princess* which is heavily skewed towards elderly tourists and is thus unlikely to be representative. Recent estimates are higher, including 50% (early data for Iceland) and as high as 86% (China). None of these studies are for representative random samples in the U.S. Based on this limited evidence, I adopt two values of the asymptomatic rate, 0.30 (for example, used here) and 0.86.

Policy paths. Shutdown policy operates through β. I consider two social distancing/economic shutdown cases. Neither are optimized and numerical values should not be taken literally. Rather, the point is to illustrate the sensitivity of the outcomes to the asymptomatic rate or, equivalently, to illustrate how different the paths for β (or equivalently, $R_0 = \beta/\gamma$) need to be to achieve a given infection rate under different values of the asymptomatic rate.

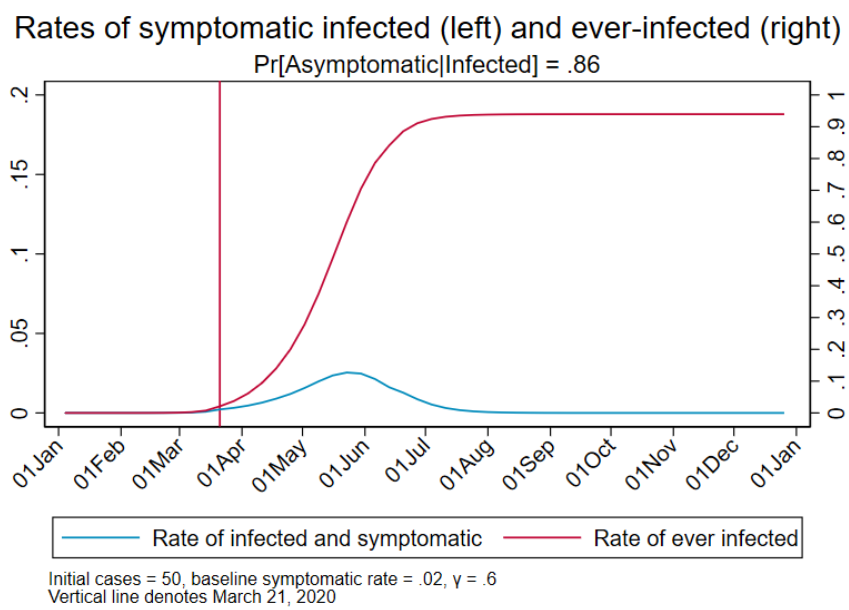
The two shutdown cases are shown in Figure 1. The first posits that moderate shutdowns continue through the end of April, then are slowly lifted, with complete lifting in the final week of June. The second posits more severe shutdowns for three months, which are then slowly lifted over the next five months. (The no-policy values of β are computed for $\pi_a = 0.3$.) Although I make no attempt to quantify the economic impacts of the two paths, they would surely be much greater for the longer, more severe path than for the early-lifting path.

Figure 1. Two policy-induced paths of R_0



The epidemiological outcomes under the two paths can now be computed. Figure 2 shows the best of the scenarios, in which the asymptomatic rate is high and the less severe path is followed. The result is the "flattening the curve" outcome, with the maximum rate of individuals who are infected and symptomatic peaking at 2.5% in the end of May. Not all these illnesses would require hospitalization.

Figure 2. High asymptomatic rate, short-duration policy



In contrast, if this policy were pursued but the asymptomatic rate is low (0.3), the result would be an overwhelming of the health care system, with rates of those infected and symptomatic approaching 20% by mid-June (Figure 3). With this low asymptomatic rate, even the draconian path for R_0 would lead to a

prolonged period of stress for the health care system (Figure 4), with the rate of those infected and symptomatic exceeding 5% in early August.

Figure 3. Low asymptomatic rate, short-duration policy

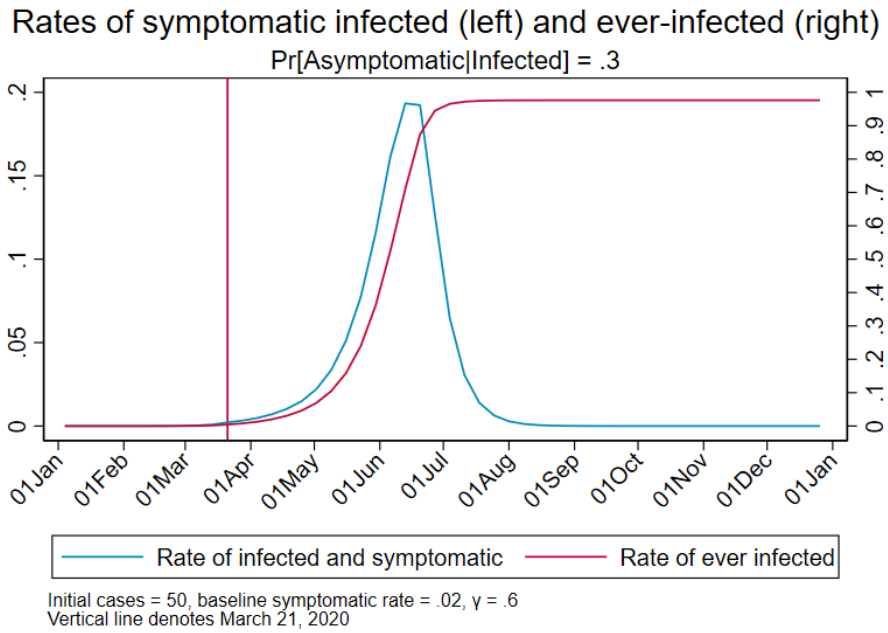
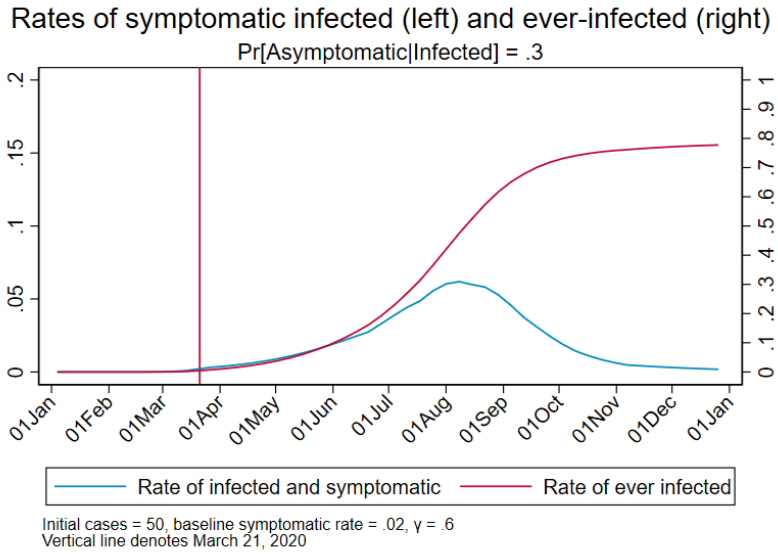


Figure 4. Low asymptomatic rate, severe long-duration policy



Summary

The most important conclusion from this exercise is that policy hinges critically on a key unknown parameter, the fraction of infected who are asymptomatic. Evidence on this parameter is scanty, however it could readily be estimated by randomized testing.

From a policy design perspective, this simplified model has the virtue of summarizing the epidemiological effect of shutdown policies in a single parameter, the contagion parameter β. In this simple model, different policies that yield the same β will have the same health outcomes. However, different policies might have very different economic costs. Thus, one way to frame the economics of shutdown policies is as finding the most efficient policies to achieve a given β, then solving for the optimal path of β that trades off the economic cost against the cost of excess lives lost by overwhelming the health care system.

References

Allen, L.J.S. (2017). "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis." *Infectious Disease Modeling* 2(2), 128-142.

Li, R. et. al. (2020)., "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)," *Science*, published online March 16, 2020 DOI: 10.1126/science.abb3221.

Qui, J. (2020). "Covert coronavirus infections could be seeding new outbreaks," *Nature* News, posted online March 20,2020 at https://www.nature.com/articles/d41586-020-00822-x.

Roser, M., H. Ritchie, and E. Ortiz-Ospina (2020). "Coronavirus Disease (COVID-19) Statistics and Research," *Our World in Data* at https://ourworldindata.org/coronavirus.