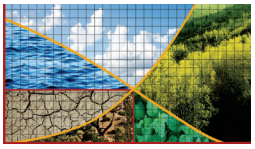


Harvard Environmental Economics Program

DEVELOPING INNOVATIVE ANSWERS TO TODAY'S COMPLEX ENVIRONMENTAL CHALLENGES

March 2020
Discussion Paper 20-82

Data Gaps and the Policy Response to the Novel Coronavirus

James H. Stock
Harvard University

HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

heep@harvard.edu
http://heep.hks.harvard.edu

Data Gaps and the Policy Response to the Novel Coronavirus

James H. Stock
Department of Economics, Harvard University
Cambridge, Massachusetts 02138

Prepared for the
Harvard Environmental Economics Program:
Discussion Paper Series
March 24, 2020

The Harvard Environmental Economics Program

The Harvard Environmental Economics Program (HEEP) develops innovative answers to today's complex environmental issues, by providing a venue to bring together faculty and graduate students from across Harvard University engaged in research, teaching, and outreach in environmental, energy, and natural resource economics and related public policy. The program sponsors research projects, convenes workshops, and supports graduate education to further understanding of critical issues in environmental economics and policy around the world. For more information, see the Program's website: http://heep.hks.harvard.edu.

Acknowledgements

The Enel Endowment for Environmental Economics at Harvard University provides major support for HEEP. The Endowment was established in February 2007 by a generous capital gift from the Enel Group, a multinational energy company and one of the world's leading integrated electricity and gas operators. HEEP also receives major support from the affiliated Enel Foundation and additional support from BP and Chevron Services Company.

HEEP enjoys an institutional home in and support from the Mossavar-Rahmani Center for Business and Government at Harvard Kennedy School. HEEP collaborates closely with the Harvard University Center for the Environment (HUCE). The Center has provided generous material support, and a number of HUCE's Environmental Fellows and Visiting Scholars have made intellectual contributions to HEEP.

HEEP and the closely-affiliated Harvard Project on Climate Agreements are grateful for additional support from the Harvard University Climate Change Solutions Fund, the Harvard Global Institute, and the Belfer Center for Science and International Affairs at Harvard Kennedy School.

Previous sponsors include: the Ash Center for Democratic Governance and Innovation at Harvard Kennedy School, Alfred P. Sloan Foundation, the AVINA Foundation, Bank of America, Castleton Commodities International LLC, ClimateWorks Foundation, Doris Duke Charitable Foundation, Duke Energy Corporation, the International Emissions Trading Association (IETA), Christopher P. Kaneb (Harvard AB 1990), the James M. and Cathleen D. Stone Foundation, the National Science Foundation, the U.S. Department of Energy, and the U.S. Environmental Protection Agency.

Citation Information

Stock, James H. "Data Gaps and the Policy Response to the Novel Coronavirus." Discussion Paper 2020-82. Cambridge, Mass.: Harvard Environmental Economics Program, March 2020.

The views expressed in the Harvard Environmental Economics Program Discussion Paper Series are those of the author(s) and do not necessarily reflect those of Harvard Kennedy School or of Harvard University. Discussion Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

Data Gaps and the Policy Response to the Novel Coronavirus

James H. Stock
Department of Economics, Harvard University

Abstract

This note lays out the basic Susceptible-Infected-Recovered (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. A key parameter, the asymptomatic rate (the fraction of the infected that are not tested under current guidelines), is not well estimated in the literature because tests for the coronavirus have been targeted at the sick and vulnerable, however it could be estimated by random sampling of the population. In this simple model, different policies that yield the same transmission rate β have the same health outcomes but can have very different economic costs. Thus, one way to frame the economics of shutdown policy is as finding the most efficient policies to achieve a given β, then determining the path of β that trades off the economic cost against the cost of excess lives lost by overwhelming the health care system.

Key words: Epidemiological models, macroeconomics, shutdowns

Acknowledgments

I thank Andy Atkeson, Oleg Itshoki, Erin Lake, and Arash Nekoei for helpful comments.

This note lays out the basic Susceptible-Infected-Recovered (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.[1] The model is calibrated to the most recent data. Its 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. First, 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 model serves to focus attention on these questions of central importance in the interaction between health policy and economic policy.

Second, 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.

Third, 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 point estimates in the epidemiological literature range from 0.18 to 0.86 (wider if sampling uncertainty is incorporated). However, the asymptomatic rate could be estimated accurately and quickly by testing a random sample of the overall population.

[1] Subsequent to writing this note, two very good and closely related papers have come out as NBER working papers, Atkeson (2020) and Eichenbaum, Rebelo, and Trabant (2020). Atkeson (2020) works through the SIR model and provides simulations for a calibrated version under different isolation scenarios. Eichenbaum, Rebelo, and Trabant (2020) merge the SIR model with a representative agent macro model to track macroeconomic outcomes. Relative to those papers, the contribution here is to illustrate the dependence of the economic outcomes on some parameters that are very poorly determined in the current literature, and to show how the model can be calibrated using data on testing from a sample that is selected under historical testing guidelines.

Fourth, the policy response and its economic consequences hinge on the value of the asymptomatic rate. We illustrate this for three policy paths for β representing various levels of shutdown and follow-up measures: two "flatten-the-curve" paths and a virus containment path. These different paths would have very different economic consequences, although those are not modeled. The health outcomes on the flatten-the-curve measures depend strongly on the asymptomatic rate.

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, which follows from the identity $N = S_t + I_t + R_t$.

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.

In this model, policy operates by manipulating the values of the parameters. The baseline values of β and γ can be considered no-policy values. Self-quarantine, social distancing, and school and 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. Policies that decrease β and/or increase γ serve to reduce R_0.

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 undetected, which different concepts although I treat them here as 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 (given γ), or alternatively could be obtained by ongoing random sampling of tests for serum antibodies in response to the coronavirus, however such tests are not yet widely available and have not been deployed on random samples in the United States.

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 as follows. 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,

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

where s_0 is the baseline rate of symptoms among the susceptible and recovered (normal colds and allergies).

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, although that should be possible. Instead, I illustrate its use and the policy importance of the key parameter π_a, the asymptomatic rate in a calibrated simulation.

Model calibration and simulation.

I now turn to a numerical illustration of the model and policy interventions. For γ, I assume that the half-life of an infection is 6 days ($\gamma = 0.55$). I set $s_0 = 0.02$. For the week of March 21, 2020, the positive testing rate in the United States was 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 solve for β such that the positive testing rate is 10% for the week ending March 21, 2020.

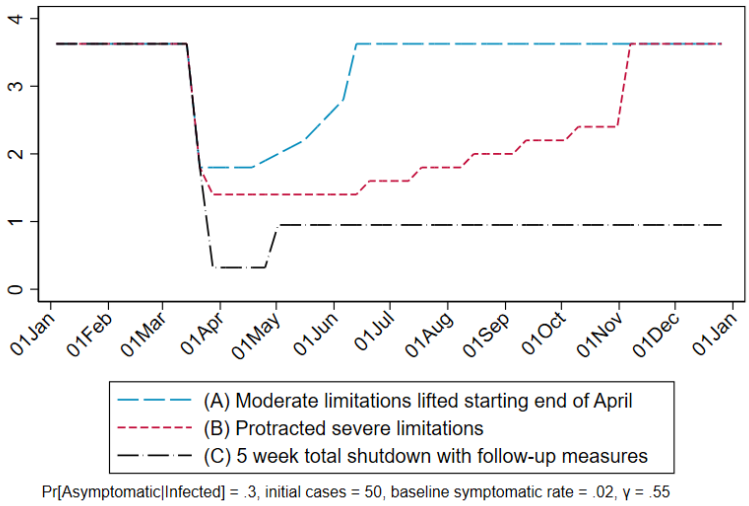
The limited available evidence on the asymptomatic rate π_a has been reviewed in Qui (March 20, 2020). Mizumoto et al's (March 12, 2020) estimate of the asymptomatic rate suggested it could be as low as 18%, however that study used data from the *Diamond Princess* which is heavily skewed towards elderly tourists and is thus unlikely to be representative. Other estimates are higher, including 31% (Nishiura et. al. (February 13, 2020)) for 565 Japanese nationals evacuated from Wuhan, however asymptomatic was defined as showing no symptoms, not simply falling short of US testing guidelines. Early data from Iceland suggest an asymptomatic rate of roughly one-half, however that uses a narrow definition of asymptomatic; a looser definition closer to US testing guidelines suggests a much higher asymptomatic rate[2]. Li et. al. (March 16, 2020) estimate a much higher rate of 86% for undetected cases for China. None of these studies are for representative random samples in the United States. Based on this limited evidence, I adopt two values of the asymptomatic rate, 0.30 (for example, used in Pueyo (March 19, 2020) and 0.86.

Policy paths. Shutdown policy operates through β. I consider three social distancing/economic shutdown cases. None 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.

[2] Iceland started a large-scale voluntary testing program in the second week of March. A March 15 report on initial results states "Of those samples which have thus far been taken, 700 have been tested. Kári [CEO of deCODE Genetics, the company running the testing] says that about half of those who tested positive have shown no symptoms, and the other half show symptoms have having a regular cold." (The Reykjuavik Grapevine, March 15, 2020 at https://grapevine.is/news/2020/03/15/first-results-of-general-population-screening-about-1-of-icelanders-with-coronavirus/, also see https://www.government.is/news/article/2020/03/15/Large-scale-testing-of-general-population-in-Iceland-underway/.) Under current US testing guidelines, both categories of cases would largely be untested and thus undetected.

The three shutdown paths of R_0 are shown in Figure 1.[3] Paths (A) and (B) represent two different "flatten the curve" strategies, in which the infection rate is managed so as not to overwhelm the health system but eventually the population achieves herd immunity. In contrast, path (C) aims to suppress the virus until a vaccine is developed. Specifically, path (A) posits moderate shutdowns that continue through the end of April, which are then slowly lifted, with complete lifting in the final week of June. Path (B) posits more severe shutdowns continuing for three months, which are then slowly lifted over the next five months. Path (C) shows what is in effect a total shutdown lasting for three months, with subsequent monitoring, ongoing testing, ongoing social distancing, and contact tracing with quarantining. Under path (C), R_0 is 0.32 for 5 weeks, a value taken from Wang et. al.'s (March 6, 2020) estimate of R_0 in Wuhan after total shutdown measures were imposed (they estimate a pre-policy R_0 of 3.86).

Figure 1. Three policy-induced paths of R_0



The epidemiological outcomes under the three paths are shown in Figures 2-4 for policy paths (A), (B), and (C) respectively. In each figure, the left panel shows the rates of those currently infected and symptomatic (in the notation above, Pr[$Symptomatic_t$, I_t]) and of the ever-infected (currently infected + recovered) for the low value of the asymptomatic rate (0.3), and the right panel shows these paths for the high value (0.86).

The outcomes for the two "flattening the curve" scenarios depend strongly on the (unknown) asymptomatic rate. The least restrictive policy (A) flattens the curve effectively if the asymptomatic rate

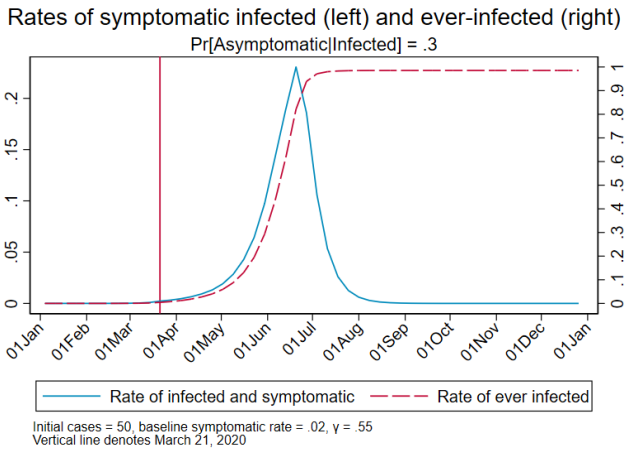
[3] The no-policy values of R_0 in these figures is computed for π_a = 0.3.

is high, and herd immunity is achieved my mid-summer. If the asymptomatic rate is low, the least restrictive policy results in very high infected symptomatic rates that would overwhelm the health system. The protracted severe limitations policy (B) flattens the curve under the lower asymptomatic rate, although the infected symptomatic rate reaches 5% in late summer. Although I do not model the economic costs, the economic costs of paths (A) and (B) are likely to be very different, with (B) resulting in high costs and a longer deeper recession.

The left panel of Figure 2 and the right panel of Figure 3 represent two high-cost scenarios, the former costly in deaths, the latter costly in economic outcomes; avoiding either requires knowledge of the asymptomatic rate and the rates of infections and recoveries.

The alternative path (C), essentially a total 5-week shutdown of face-to-face interaction as was done in Wuhan, would have the most severe immediate economic costs. Under the parameters here, this suffices to suppress the virus. If R_0 is kept below 1 until a vaccine is developed, then the total rate of ever-infected in the population remains low.

Figure 2. Policy path (A) with low (left) and high (right) asymptomatic rates



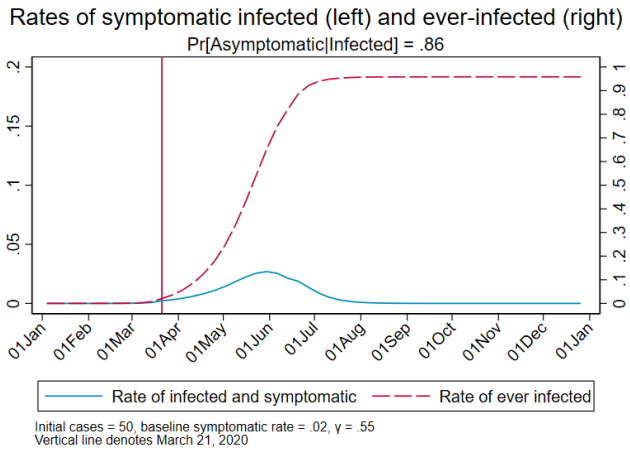
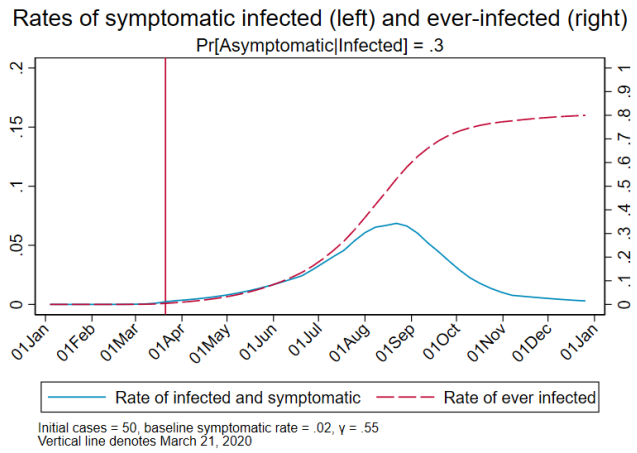


Figure 3. Policy path (B) with low (left) and high (right) asymptomatic rates



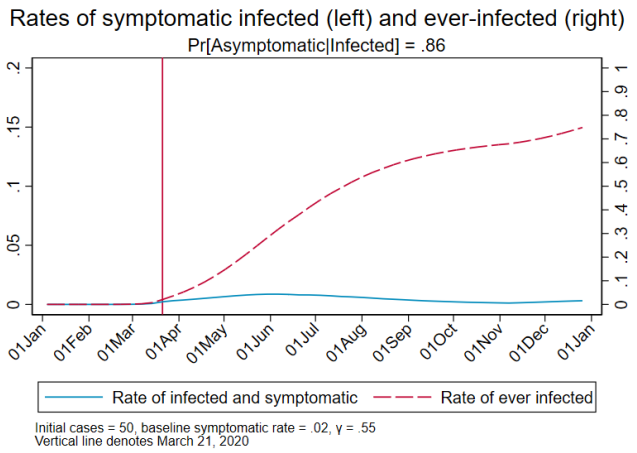
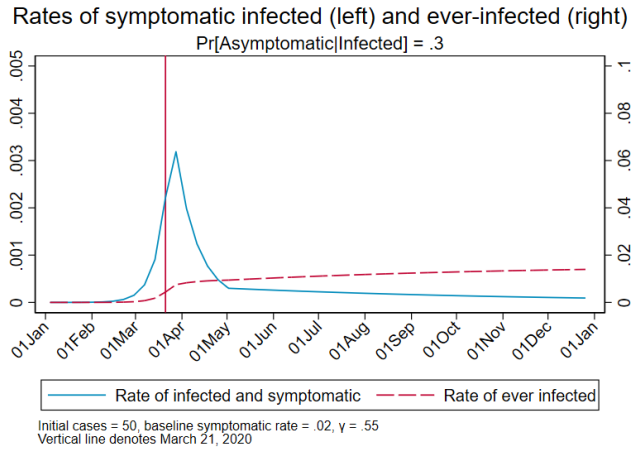
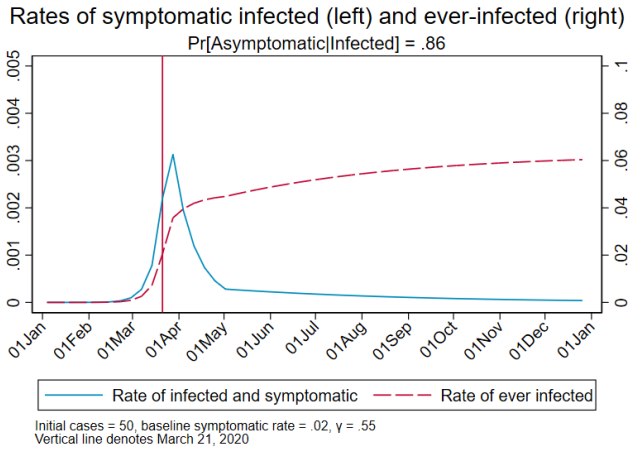


Figure 4. Policy path (C) with low (left) and high (right) asymptomatic rates

Note: Axis values differ from Figures 2 and 3.





Summary

Policy outcomes hinge critically on a key unknown parameter, the fraction of infected who are asymptomatic, and on the current rates of infected and recovered in the population. Evidence on the asymptomatic rate is scanty, however it could readily be estimated by randomized testing.

From an economic 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.

Atkeson, A. (2020), "What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios," NBER Working Paper 26867, March 2020.

Eichenbaum, M., S. Rebelo, and M. Trabant (2020), "The macroeconomics of epidemics," NBER Working Paper 26882, March 2020.

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 at https://science.sciencemag.org/content/early/2020/03/13/science.abb3221.

Mizumoto, K. et. al., "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020," *Eurosurveillance* 25(10), March 12, 2020 at https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180.

Nishiura, H. et. al. (2020), "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)," forthcoming, *International Journal of Infectious Disease*, posted online February 13, 2020 at https://doi.org/10.1016/j.ijid.2020.03.020.

Pueyo, T. (2020), "Coronavirus: The Hammer and the Dance," *Medium* March 19, 2020 at https://medium.com/@tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56

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.

Wang, C. et. al. (2020), “Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China,” March 6, 2020 at https://www.medrxiv.org/content/10.1101/2020.03.03.20030593v1.