# Voluntary and Mandatory Social Distancing: Evidence on COVID-19 Exposure Rates from Chinese Provinces and Selected Countries

*Alexander Chudik, M. Hashem Pesaran, Alessandro Rebucci*

# Voluntary and Mandatory Social Distancing: Evidence on COVID-19 Exposure Rates from Chinese Provinces and Selected Countries

## Abstract

This paper considers a modification of the standard Susceptible-Infected-Recovered (SIR) model of epidemic that allows for different degrees of compulsory as well as voluntary social distancing. It is shown that the fraction of population that self-isolates varies with the perceived probability of contracting the disease. Implications of social distancing both on the epidemic and recession curves are investigated and their trade-off is simulated under a number of different social distancing and economic participation scenarios. We show that mandating social distancing is very effective at flattening the epidemic curve, but is costly in terms of employment loss. However, if targeted towards individuals most likely to spread the infection, the employment loss can be somewhat reduced. We also show that voluntary self-isolation driven by individual's perceived risk of becoming infected kicks in only towards the peak of the epidemic and has little or no impact on flattening the epidemic curve. Using available statistics and correcting for measurement errors, we estimate the rate of exposure to COVID-19 for 21 Chinese provinces and a selected number of countries. The exposure rates are generally small, but vary considerably between Hubei and other Chinese provinces as well as across countries. Strikingly, the exposure rate in Hubei province is around 40 times larger than the rates for other Chinese provinces, with the exposure rates for some European countries being 3-5 times larger than Hubei (the epicenter of the epidemic). The paper also provides country-specific estimates of the recovery rate, showing it to be about 21 days (a week longer than the 14 days typically assumed), and relatively homogeneous across Chinese provinces and for a selected number of countries.

*Alexander Chudik*
*Federal Reserve Bank of Dallas / USA*
*alexander.chudik@dal.frb.org*

*M. Hashem Pesaran*
*University of Southern California / USA*
*pesaran@usc.edu*

*Alessandro Rebucci*
*Johns Hopkins University Carey Business School*
*arebucci@jhu.edu*

# 1 Introduction

The COVID-19 pandemic has already claimed many lives and is causing an unprecedented and widespread disruption to the world economy. China responded to the initial outbreak with draconian social distancing policies which are shown to be effective in containing the epidemic, but at the cost of large short term losses in employment and output. Other countries have responded more timidly, either by deliberate choice, as in the United States, or due to implementation constraints, as in some European countries. The purpose of this paper is to evaluate the impact of alternative mitigation or containment policies on both the epidemic and the so-called recession curves, and to empirically compare their implementation across countries.

Most importantly we consider both government-mandated social distancing policies, and voluntary self-isolation, and endogenize the fraction of the population that remain exposed to the virus within a standard Susceptible-Infected-Recovered model (SIR). Specifically, we distinguish between individuals exposed to COVID-19 and those isolated from the epidemic. We decompose the population, $P$, into two categories: those who are exposed to COVID-19 in the sense that they can contract the virus because they are not isolated and they have not been infected yet, $P_E$; and the rest, $P_I$, that are isolated and therefore taken out of harm's way. We denote the strength of the mitigation policy by $1 - \lambda$, where $\lambda$ is the proportion of population that is exposed to COVID-19, defined as $\lambda = P_E/P$. Initially we focus on the relatively simple case where $\lambda$ is set at the outset of the spread of the epidemic, close to what we believe China did after the start of the epidemic in Wuhan. We also consider a variation of the SIR model where $\lambda$ changes due to the voluntary decision to isolate at the micro level. Using a simple decision model we show that the proportion of the population that self-isolates rises with the probability of contracting the disease. We approximate this probability with the number of active cases and show (by simulation) that the effect of self-isolation occurs as the epidemic nears its peak, and is relatively unimportant during the early or late stages of the epidemic. A coordinated social policy is required from the early outset of the epidemic to flatten the epidemic curve.

We then model the short-term impact of the epidemic on employment. This permits an eval-

uation of the costs and benefits of alternative societal decisions on the degree and the nature of government-mandated containment policies by considering alternative values of $\lambda$ in conjunction with an employment loss elasticity, $\alpha$, that allows a given social distancing policy to have different employment consequences. In the extreme case where the incidence of social distancing is uniform across all individuals and sectors, a fall in $\lambda$ results in a proportionate fall in employment, and $\alpha = 1$. But by enabling individuals to isolate and to work from home, together with wide spread targeted testing for the virus plus the use of protective clothing and equipment, it is possible to mitigate somewhat the economic costs of social distancing policies. We simulate the employment loss for alternative values of $\lambda$ and $\alpha$ and find that, for sufficiently low values of $\lambda$ required to manage the peak of hospitalization and death from COIVD-19, the economic costs could be substantial even with smart social distancing policies. We also simulate the duration of the epidemic to be around 120 days, with a sizeable part of the employment loss occurring close to the peak of the epidemic.

Whilst there is ample medical and biological evidence on the key parameters of the SIR model, namely the basic reproduction rate, $R_0$, and the recovery rate, $\gamma$, to our knowledge there are no direct estimates of $\lambda$. A recent report from the Imperial College COIVD-19 Response Team uses a Bayesian hierarchical model to infer the impact of social distancing policies implemented across 11 European countries, see Flaxman et al. (2020). They use the number of observed deaths to infer the number of infections and do not make use of confirmed infections that are subject to significant measurement errors due to limited testing. Whilst acknowledging the measurement problems, in this paper we provide estimates of $\gamma$ and $\lambda$ using daily data on confirmed, recovered and death cases from the Johns Hopkins University (JHU) hub.[1] Using a discretized version of our modified SIR model we derive reduced form regressions in confirmed recoveries and the number of active cases that allow for systematic and random measurement errors. We show that $\gamma$ can be identified assuming that confirmed infected and recovery cases are subject to a similar degree of mis-measurement. We also show that, for a given value of $R_0$, the social distancing parameter, $\lambda$, can be identified up to a fraction which is determined by the scale of mis-measurement of reported

---

[1] Available at https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data.

active cases. We calibrate this fraction using the data from the Diamond Princess cruise ship reported by Moriarty et al. (2020).

We first use daily data on Chinese provinces with complete history of the course of the epidemic. The estimates of the recovery rates are very similar across the Chinese provinces and lie in the range of 0.033 (for Beijing) and 0.066 (for Hebei). We also find that the random measurement in the underlying data is relatively unimportant for the estimation of $\gamma$. The mean estimate of $\gamma$ across the Chinese provinces is around 0.046 which corresponds to around 22 days from infection to recovery (or death). This estimate is substantially larger than the 14 days typically assumed in designing quarantine policies. Setting $\gamma = 0.046$ and $R_0$, we then proceed to estimate $\lambda$ (up to the scaling fraction). We find that for Chinese provinces $\lambda$ is very small even if we allow for a significant under-recording of infected and recovered cases. We find that, with the exception of Hubei province (the epicenter of the epidemic), the share of exposed population across other provinces was less than 1 individuals per 100,000! This is an astonishingly low rate and is consistent with dramatically falling estimates of the effective reproduction rate at the onset of the epidemic in China. In contrast, the estimates of $\lambda$ which we have obtained for European countries are significantly higher even when compared to the relatively high exposure rates for Hubei, with a substantial heterogeneity across countries. In particular, we estimate exposure rates for Italy and Spain to be almost five times the rate estimated for Hubei.

To summarize, *our theoretical analysis shows that voluntary social distancing is likely to be effective only when the epidemic begins to approach its peak, and mandated social distancing to flatten the curve is required from the early phases of the epidemic. Our estimates show that in order to flatten the epidemic curve very strict mandatory policies are necessary, as in the case of the Chinese provinces excluding the Hubei epicenter show. Unfortunately, our estimates suggest that, despite the time-lag in the contagion from China to other countries, an inadequate and uncoordinated policy response resulted in exposure rates outside of China that are multiples of those documented at the epicenter of the epidemic in Hubei.*

**Related Literature**

The characteristics and the economic consequences of the COVID-19 outbreak, and of policies

to contain its spread, are the subject of a fast growing body of research. Scientific evidence based on more accurate data at the local level has begun to document the rate of transmission and incubation periods. The literature has also begun to document the role of mitigation policies in reducing transmission, and the rate of asymptomatic transmission.

Kucharski et al. (2020) estimate that, in China, the effective reproductive rate $R_t$ fell from 2.35 one week before travel restrictions were imposed on Jan 23, 2020, to 1.05 one week after travel restrictions. They use a SIR model and estimate it to forecast the epidemic in China, extending the model to explicitly account for infections arriving and departing via flights. Using data from Wuhan, Wang et al. (2020) report a baseline reproductive rate of 3.86, that fell to 0.32 after the vast lock-down intervention. They also find a high rate of asymptomatic transmission.

Work on the economic impact of the epidemic is just starting, as the data are only partially available. Atkeson (2020) explores the trade-off between the severity and timing of suppression of the disease, for example through social distancing, and the progression of the disease in the population in simulations of a SIR model like ours with exposed and not exposed population, but does not provide estimates and does not focus on the share of the exposed population, nor does he provides estimates of the model parameters.

Berger, Herkenhoff, and Mongey (2020) show that testing at a higher rate in conjunction with targeted quarantine policies can reduce both the economic impact of the COIVD-19 and peak symptomatic infections. As noted above, by selectively applying social distancing policies (with different $\alpha$ parameters) it is also possible to reduce both the economic impact of the epidemic and the peak symptomatic infections. Related to this, using data on the Spanish flue, Correia, Luck, and Verner (2020) find that cities that intervened earlier and more aggressively do not perform worse and, if anything, grow faster after the pandemic is over. These findings thus indicate that containment policies not only lower mortality, they also mitigate the adverse long term economic consequences of a pandemic.

Fang, Wang, and Yang (2020) analysis of Chinese efforts to contain the COIVD-19 outbreak measures the effectiveness of the lock-down of Wuhan and enhanced social distancing policies in other cities. They produce evidence for all Chinese provinces and show that these policies con-

tributed significantly to reducing the total number of infections outside of Wuhan.

Stock (2020) focuses on measurement error and explores the benefits of randomly testing the general population to determine the asymptomatic infection rate.

Eichenbaum, Rebelo, and Trabandt (2020) discuss the trade off between the economic costs of containment policies — which could include a recession — and the number of lives saved in a model in which agents optimize and the probability of infection is endogenous. Barro, Ursua, and Weng (2020) estimate death rates and output losses based on 43 countries during the 1918-1920 Spanish flu. They find a very high death rate, with 39 million deaths, or 2.0 percent of world population, implying 150 million deaths when applied to current population. According to their estimates, the Spanish flue resulted in economic declines for GDP and consumption in the typical country of 6 and 8 percent, respectively.

Linton (2020) uses a reduced form quadratic time trend model in log of new cases and new deaths to predict the peak of COIVD-19 for a large number of countries.

As far as we are aware, no study has modelled the difference between government-mandated and self-imposed isolation and their implication for the flattening of the economic and pandemic curves that we consider in this paper.

The rest of the paper is organized as follows. Section 2 sets out the modified SIR model with social distancing. Section 3 analyzes the distinction between mandatory and voluntary isolation. Section 4 discusses the trade off between containing the epidemic and the employment losses that depend on the share of exposed population. Section 5 sets out the econometric and measurement models and reports the estimation results. Section 6 concludes.


## 2    A Discrete-time SIR Model with Mitigation Policy

There are many approaches to modelling the spread of epidemics. The basic mathematical model used by many researchers is the susceptible-infective-removed (SIR) model advanced by Kermack and McKendrick (1927). This model, and its various extensions, has been the subject of a vast number of studies, and has been used extensively over the past few months to investigate the spread

of COVID-19. A comprehensive treatment is provided by Diekmann and Heesterbeek (2000) with further contributions by Metz (1978), Satsuma et al. (2004), Harko et al. (2014), Salje et al. (2016), amongst many others.

The basic SIR model considers a given population of fixed size $P$, composed of three distinct groups, those individuals in period $t$ who have not yet contracted the disease and are therefore susceptible, denoted by $S_t$; the 'removed' individuals who can no longer contract the disease, consisting of recovered and deceased, denoted by $R_t$; and those who remain infected at time $t$ and denoted by $I_t$. Thus

$$P = S_t + I_t + R_t. \tag{1}$$

As it stands, this is an accounting identity, and it is therefore sufficient to model $S_t$ and $I_t$ and obtain $R_t$ as the remainder. The SIR model is typically cast in a set of differential equations, which we discretize and write as the following difference equations (for $t = 1, 2, ..., T$)

$$S_{t+1} - S_t = -\beta s_t I_t, \tag{2}$$

$$I_{t+1} - I_t = (\beta s_t - \gamma) I_t, \tag{3}$$

$$R_{t+1} - R_t = \gamma I_t, \tag{4}$$

where $\beta$ and $\gamma$ are the key parameters of the epidemic. $\beta$ is the rate of transmission and $\gamma$ is the recovery rate. In this model it is assumed that an infected individual in period $t$ causes $\beta s_t$ secondary infections, where $s_t = S_t/P$ is the share of susceptible individuals in the total population. The time profile of $I_t$ critically depends on the basic reproduction number, defined as the expected number of secondary cases produced by a single infected individual in a completely susceptible population, denoted by $R_0 = \beta/\gamma$. The parameter $\beta$ is determined by the biology of the virus, and is assumed to be constant over time and homogeneous across countries and regions. The recovery rate, $\gamma$, can also be written as $\gamma = 1/d$, where $d$ denotes the number of days to recover or die from the infection. We assume that $\gamma$ is constant over time, but allow it to vary across countries and regions, reflecting the differences in the capacity of the local health care systems to treat the infected population.

The epidemic begins with non-zero initial values $I_1 > 0$ and $S_1 > 0$, and without any mitigation policies in place it will spread widely if $R_0 > 1$, ending up infecting a large fraction of the population if $R_0$ is appreciably above unity. We show below that the steady state value of this proportion is given by $\pi_0 = (R_0 - 1)/R_0$. In the case of COVID-19, a number of different estimates have been suggested in the literature, placing $R_0$ somewhere in the range of 2.4 to 3.9.[2]

In the simulations, and the empirical analysis to follow, we adopt a central estimate and set $R_0 = 3$. As we shall see the SIR model predicts that in the absence of social distancing as much as 2/3 of the population could eventually become infected before the epidemic runs its course– the so-called herd immunity solution to the epidemic. Such an outcome will involve unbearable strain on national health care systems and a significant loss of life, and has initiated unparalleled mitigation policies first by China and South Korea, and more recently by Europe, US and many other countries. Such interventions, which broadly speaking we refer to as "social distancing", include case isolation, banning of mass gatherings, closures of schools and universities, and even local and national lock-downs.

To investigate the economic implications of such policies we first modify the SIR model by decomposing the total population, $P$ into two categories, those who are exposed to COVID-19 in the sense that they could catch the virus (they have not been infected yet), $P_E$, and the rest, $P_I$, who are isolated and therefore taken out of harm's way. We denote the strength of the mitigation policy by $1 - \lambda$, where $\lambda$ is the proportion of population that is exposed to COVID-19, defined as $\lambda = P_E/P$. In practice, $\lambda$ will be time-varying and most likely there will be feedbacks from the progress of the epidemic to the coverage of the intervention policies. Here we consider the relatively simple case where $\lambda$ is set at the outset of the spread of the epidemics, close to what we believe China did after the start of the epidemic in Wuhan.

In the presence of the social distancing intervention characterized by $\lambda$, the equations of the

---

[2]Using data from Wuhan, Wang et al. (2020) report a pre-intervention reproductive rate of 3.86. Kucharski et al. (2020) estimate that, in China, the reproductive rate was 2.35 one week before travel restrictions were imposed on Jan 23, 2020. Ferguson et al. (2020) made baseline assumption of $R_0 = 2.4$ based on the fits to early growth-rate of epidemic in Wuhan (and also examined values of 2.0 and 2.6) based on fits to the early growth-rate of the epidemic in Wuhan by Li et al. (2020) and Riou and Althaus (2020).

SIR model now become (noting that the population exposed to the virus is now $P_E$):

$$S_{t+1} - S_t = -\beta \left( \frac{S_t}{P_E} \right) I_t,$$

$$I_{t+1} - I_t = \left[ \beta \left( \frac{S_t}{P_E} \right) - \gamma \right] I_t,$$

$$\lambda P = S_t + I_t + R_t.$$

Dividing both sides of the above equation by $P$ and using the fractions $s_t = S_t/P$, $i_t = I_t/P$ and $r_t = R_t/P$, we have

$$s_{t+1} - s_t = -\left( \frac{\beta}{\lambda} \right) s_t i_t, \tag{5}$$

$$i_{t+1} - i_t = \left[ \left( \frac{\beta}{\lambda} \right) s_t - \gamma \right] i_t, \tag{6}$$

and

$$\lambda = s_t + i_t + r_t. \tag{7}$$

Given $\beta$, the fraction of total exposed population, $\lambda$, determines the effective transmission rate, $\theta = \beta/\lambda$. When $\lambda = 1$, the whole population is exposed, and the effective transmission rate coincides with the biological one, $\beta$.

The system equations (5) and (6) can be solved by iterating forward from some non-zero initial values, with $i_1$ a small fraction and $s_1 = \lambda - i_1$, since at the start of the epidemic we can safely assume that $r_1 = 0$. Iterating (6) forward from $i_1 > 0$, and for given values of $s_1, s_2, ...s_t$ we have (where we have replaced $\theta = \beta/\lambda = \gamma (R_0/\lambda)$)

$$i_{t+1} = \left( \prod_{\tau=1}^{t} \rho_\tau \right) i_1, \tag{8}$$

where $\rho_\tau = 1 + \gamma \left[ (R_0/\lambda) s_\tau - 1 \right]$. Initially, where few are infected and $s_\tau$ is close to $\lambda$, $\rho_\tau > 1$ and the number of infected individuals rises exponentially fast so long as $R_0 > 1$. But as the disease spreads and recovered and/or deceased are removed, at some point in time $t = t^*$, $s_\tau$ starts to fall for $\tau > t^*$ such that $\rho_\tau < 1$ from $\tau > t^*$, and eventually $lim_{t \to \infty} \left( \prod_{\tau=1}^{t} \rho_\tau \right) = 0$. Hence, $lim_{t \to \infty} (i_t) = i^* = 0$.

Further, $lim_{t\to\infty}(i_{t+1}/i_t) = 1$, and from (6) we have $lim_{t\to\infty}s_t = s^* = \lambda(\gamma/\beta) = \lambda/R_0$, and using the identity $\lambda = s^* + r^*$, we finally obtain the following expression for the total number of infected cases as a fraction of the population $(c^*)$:

$$c^* = r^* = \lambda - \lambda/R_0 = \frac{\lambda(R_0 - 1)}{R_0}. \tag{9}$$

The choice of $\lambda$ also has important implications for the steepness and and the peak of the epidemic curve, and can be used to flatten the trajectory of $i_t$. To this end, and also for the purpose of estimating $\lambda$ using data realizations from completed epidemics, we first eliminate $s_t$ from the equation for $i_t$ noting from (6) and (5) that

$$\frac{s_{t+1}}{s_t} = 1 - \theta i_t \quad \text{and} \quad \frac{i_{t+1}}{i_t} = 1 - \gamma + \theta s_t. \tag{10}$$

Since $\theta > 0$, solving for $s_t$, we have $s_t = \theta\left(\frac{i_{t+1}}{i_t} - 1 + \gamma\right)$, and hence

$$\frac{s_{t+1}}{s_t} = \frac{\left(\frac{i_{t+2}}{i_{t+1}} - 1 + \gamma\right)}{\left(\frac{i_{t+1}}{i_t} - 1 + \gamma\right)} = 1 - \theta i_t,$$

which yields the following second-order non-linear difference equation in $i_t$

$$i_{t+1} = i_t^2/i_{t-1} + \theta\left[i_t i_{t-1}(1 - \gamma) - i_t^2\right], \ t = 1, 2, ..., T, \tag{11}$$

with the initial values $i_1$ and $i_2 = (1 - \gamma + \theta s_1)i_1$, where $s_1 = \lambda - i_1$. Realizations on $i_t$, for $t = 1, 2, ..., T$, can also be used to estimate $\gamma$ and $\theta$ from the above non-linear autoregression, but it is important to note that $\beta$ and $\lambda$ cannot be separately identified without further *a priori* knowledge. In the empirical analysis that follows we set $R_0$ a priori and estimate $\lambda$ from $\gamma R_0/\theta$, as $\beta = \gamma R_0$ and $\theta = \beta/\lambda$.

As an illustration in Figure 1 we show the time profile of $i_t(\lambda)$ using the parameter values $R_0 = 3$ and $\gamma = 1/d = 1/14$ and the initial values $i_1(\lambda) = \lambda/1000$, and $i_2(\lambda) = [1 - \gamma + (\gamma R_0/\lambda)s_1(\lambda)]i_1(\lambda)$, where $s_1(\lambda) = \lambda - i_1$. Consider the following social distancing coefficients, $\lambda = 1, 0.75$, and $0.50$.

The time profiles of the infected (as the fraction of population) peak after 52 days and does not seem to depend much on the choice of $\lambda$. But the choice of $\lambda$ is clearly important for flattening the curve and reduces the peak of infected from 31 percent when $\lambda = 1$, to 23 per cent for $\lambda = 0.75$, and to 15 per cent for $\lambda = 0.50$.

**Figure 1**: Simulated values of $i_t(\lambda)$ for different social distancing coefficients: $\lambda = 1$ (blue), 0.75 (red), and 0.5 (green)
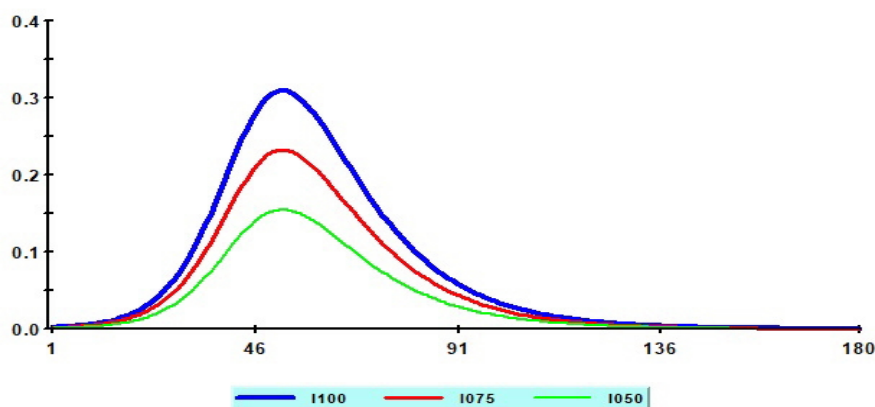


## 3 A voluntary model of social isolation: case of time-varying $\lambda$

So far we have assumed that $\lambda$, the proportion of population that can be infected is fixed and set exogenously by central authorities. In practice, the degree of social distancing also depends on the extent to which individuals follow the rules, which could depend on the fear of contracting the disease and most likely will depend on the number of those who are already infected, and an individual's perception of the severity of the epidemic and its rate of spread. Even central authorities can be slow to respond when the number of active cases is small and they might delay or start with a low level of social distancing and then begin to raise it as the number of infected cases start to increase rapidly. In the context of our modified SIR model we can allow for such time variations in $\lambda$ by relating the extent of social isolation in day $t$, measured by $1 - \lambda_t$, to the probability of contracting the disease.

10

More formally, consider an individual $j$ from a fixed population of size $P$ in the day $t$ from the start of an epidemic, and suppose the individual in question is faced with the voluntary decision of whether to isolate or not. Under self-isolation the individual incurs the loss of wages net of transfers, amounting to $(1 - \tau_j)w_j$, plus the inconvenience cost, $a_j$, of being isolated. For those individuals who can work from home $\tau_j$ is likely to be 1 or very close to it. But for many workers who are furloughed or become unemployed, $\tau_j$ is likely to be close to zero, unless they are compensated by transfers from the government. On the other hand, if the individual decides not to self-isolate then he/she receives the uncertain pay-off of $(1 - d_{jt})w_j - d_{jt}\phi_j$, where $d_{jt}$ is an indicator which takes the value of unity if the individual contracts the disease and zero otherwise. $\phi_j$ represents the cost of contracting the disease and is expected to be quite high. We are ruling out the possibility of death as an outcome. In this setting the individual decides to self-isolate if the sure loss of self-isolating is less than the expected loss of not self-isolating, namely if

$$(1 - \tau_j)\, w_j + a_j < E\left[d_{jt}\phi_j - (1 - d_{jt})w_j \,|\mathcal{I}_{t-1}\right], \tag{12}$$

where $\mathcal{I}_{t-1}$ is the publicly available information that includes $i_{t-1}$, the proportion of population being infected in day $t-1$. We assume that the probability of anyone contracting the disease is uniform across the population and this is correctly perceived to be given by $\pi_{t-1}$. Hence $E\left(d_{jt}\,|\mathcal{I}_{t-1}\right) = \pi_{t-1}$, and the condition for self-isolating can be written as

$$(2 - \tau_j)w_j + a_j < \pi_{t-1}(w_j + \phi_j),$$

or as

$$\frac{2 - \tau_j + (a_j/w_j)}{1 + (\phi_j/w_j)} = \mu_j < \pi_{t-1}. \tag{13}$$

Since $\pi_{t-1} \leq 1$, then for individual $i$ to self-isolate we must have $\mu_j < 1$, (note that $\mu_j \geq 0$, with $\mu_j = 0$ when $\phi_j \to \infty$) or if

$$\phi_j/w_j > a_j/w_j + (1 - \tau_j). \tag{14}$$

Namely, if the relative cost of contracting the disease, $\phi_j/w_j$ is higher than the inconvenience cost

of self-isolating plus the proportion of wages being lost due to self-isolation. Also, an individual is more likely to self-isolate voluntarily if the wage loss, measured by $\tau_j$, is low thus providing an additional theoretical argument in favor of compensating some workers for the loss of their wages, not only to maintain aggregate demand but to encourage a larger fraction of the population to self-isolate.

The above formulation also captures the differential incentive to self-isolate across different age groups and sectors of economic activity. Given that the epidemic affects the young and the old differently, with the old being more at risk as compared to the young, then $\phi_{old} > \phi_{young}$, and the old are more likely to self-isolate. Similarly, low-wage earners are more likely to self-isolate as compared to high-wage earners with the same preferences ($\phi_j$ and $\alpha_j$), and facing the same transfer rates, $\tau_j$. But the reverse outcome could occur if low-wage earner face a higher rate of transfer as compared to the high-wage earners. These and many other micro predictions of the theory can be tested. But here we are interested in the aggregate outcomes, in particular the fraction of the population that voluntarily self-isolates.

Denote the fraction of the population in day $t$ who are self-isolating voluntarily by $v_t(P)$ and using (13) note that

$$v_t(P) = P^{-1} \sum_{j=1}^{P} I\left(\mu_j < \pi_{t-1}\right).$$

Suppose now that condition (14) is met and $0 \leq \mu_j < 1$. Further suppose that the differences in $\mu_j$ across $j$ can be represented by a continuous distribution function, $F_\mu(.)$. Then assuming that $\mu_j$ are independently distributed across $j$, by the standard law of large numbers we have

$$v_t = lim_{P \to \infty}([v_t(P)] = Pr\left(\mu_j < \pi_{t-1}\right) = F_\mu\left(\pi_{t-1}\right). \tag{15}$$

In practice, although $P$ is fixed, it is nevertheless sufficiently large (in millions) and the above result holds, almost surely.[3]

In the case where a fixed fraction, $1 - \lambda$, of the population are placed under compulsory social

---

[3]This limiting result holds even if $\mu_j$ are cross correlated so long as the degree of cross correlation across $\mu_j$ is sufficiently weak.

distancing, and the remaining $\lambda$ fraction of population decides to self-isolate voluntarily, the overall fraction of the population that isolates either compulsory or voluntarily is given by

$$1 - \lambda_t = (1 - \lambda) + \lambda F_\mu (\pi_{t-1}),$$

which yields the following expression for the fraction of population in day $t$ that is not self-isolating

$$\lambda_t = \lambda \left[1 - F_\mu (\pi_{t-1})\right]. \tag{16}$$

Assuming that $\mu_j$ is distributed uniformly over $0 \leq \mu_j < 1$, we have $\lambda_t = \lambda(1 - \pi_{t-1})$. Other distributions, such as Beta distribution can also be considered. But, as to be expected, it is clear that $\lambda_t$ is inversely related to $\pi_{t-1}$. The higher the probability of contracting the disease the lower the fraction of the population that will be exposed to the disease.

In order to integrate the possibility of time variations in $\lambda$ to the SIR model, we need to provide an approximate model for $\pi_{t-1}$, noting that $\pi_{t-1}$ is not the true probability of contracting the disease (which itself depends on $\lambda_t$ in a circular manner), but the subjective (or perceived) probability by individuals. As a simple, yet plausible approximation, we suppose that $\pi_{t-1} = \kappa i_{t-1}$, where $\kappa > 0$, and $\kappa \sup_t(i_t) < 1$, and write the modified SIR model as

$$s_{t+1} - s_t = -\left(\frac{\beta}{\lambda_t}\right) s_t i_t, \tag{17}$$

$$i_{t+1} - i_t = \left[\left(\frac{\beta}{\lambda_t}\right) s_t - \gamma\right] i_t, \tag{18}$$

$$\lambda_t = \lambda(1 - \kappa i_{t-1}), \tag{19}$$

which can be solved iteratively from the initial values $i_1$ and $s_1$. This formulation clearly reduces to the time-invariant case when $\kappa = 0$. Since $i_{t-1} \geq 0$, then $\lambda_t \leq \lambda$ and the proportion of the population who are in harm's way declines as the epidemic spreads, and rises towards $\lambda$, as the epidemic starts to wane. Following a similar line of reasoning as before, it is easily established that $i^* = lim_{t\to\infty}(i_t) = 0$, and $\lambda^* = lim_{t\to\infty}\lambda_t = \lambda$, with the rest of the results for the case of fixed $\lambda$ holding in the limit.

Figure 2 below shows the simulated values of $\lambda_t$ from iterating equations (17), (18) and (19) forward with parameters $R_0 = 3, \gamma = 1/14, \lambda = 0.5$ and $\kappa = 1.5$. The differences in the time profiles of $i_t$ without feedback effects ($\lambda = 0.5$, and $\kappa = 0$) and the ones with feedback effects ($\lambda = 0.5$, and $\kappa = 1.5$) are shown in Figure 3. As can be seen, by relating $\lambda_t$ to vary inversely with $i_{t-1}$, it is possible to flatten the peak of infected cases curve, and reduce the adverse public health and economic implications of the epidemic. But voluntary social distancing starts to have an effect only once the epidemic is already widely spread, and some coordinated social policy is clearly needed from the out-set, and before the epidemic begins to spread widely.

**Figure 2**: Simulated values of $\lambda_t$ in the case of the SIR model with parameters

$R_0 = 3, \gamma = 1/14, \lambda = 0.5$ and $\kappa = 1.5$.
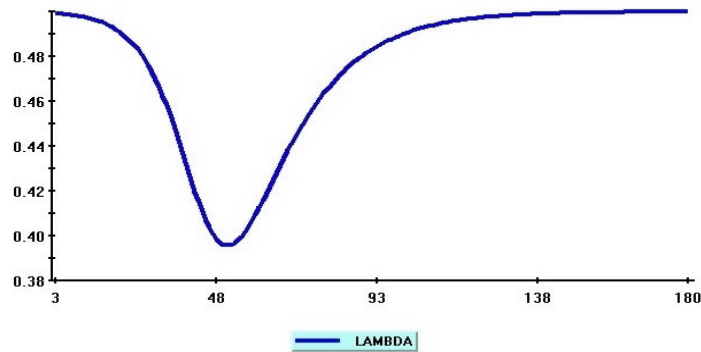


**Figure 3**: Time profiles of $i_t$ with a fixed $\lambda = 0.5$ and time-varying lambda with $\kappa = 0$ (blue) and

$\kappa = 1.5$ (red)



14
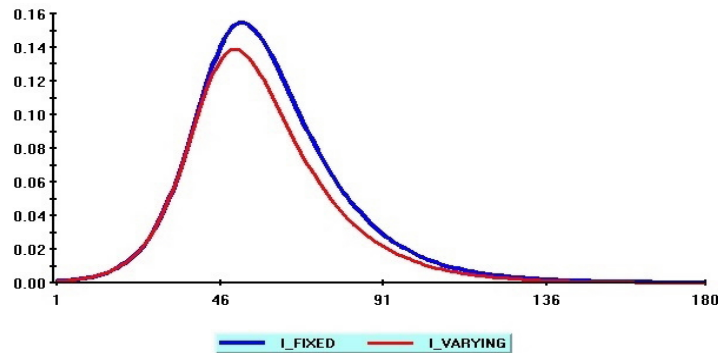
# 4  The Economic Cost of Mitigating Epidemics

The choice of $\lambda$ plays a critical role in establishing a balance between the height of the infection and the associated economic costs. The reduction in $\lambda$ can be achieved through social distancing. There is a clear trade off between the adverse effects of reducing $\lambda$ on the employment rate and its positive impact on reducing the fraction of the population infected, $i_t(\lambda)$, and hence removed from the work force. There is also the further trade off between the employment rate and the death rate due to the spread of the epidemic for different choices of $\lambda$. However, we do not model the death rate or take into account its economic or social cost.

Since the duration of the epidemic is expected to be relatively short, 3 to 4 months at most, it is reasonable to assume that the immediate economic impact of the epidemic will be on the rate of employment. In the absence of the epidemic, we assume that the rate of employment in day $t$ is given by $e_t = E_t/P$ where $E_t$ is the counterfactual level of employment during day $t$, and $P$ is the population taken as given. For the US economy the current value of $e_t$ is around 60%, which we take to be the counterfactual employment rate.

Consider now the rate of employment during the spread of the epidemic, $t = 1, 2, ..., T$, under the social distancing policy $\lambda$ ($0 < \lambda \leq 1$). The effect of the epidemic on the employment level is two-fold. First it reduces the number in employment directly by $f(\lambda)$, where $f(1) = 1$, and $f(\lambda)$ is an increasing function of $\lambda$, with $f''(\lambda) \leq 0$, for $\lambda$ in the range $0 < \lambda \leq 1$. In the extreme scenario where the incidence of social distancing is uniform across all individuals and all sectors of the economy, we have $f(\lambda) = \lambda$. But in practice the fall in employment is likely to be less than proportional since some who work from home are less affected by social distancing as compared to those who are fired because of down-sizing and firm closures. It is also possible to mitigate the negative employment effects of social distancing by focussing on sectors of the economy that are less affected by social distancing, by embarking on intensive and targeted testing, contract tracing, and by more extensive use of protective clothing and equipment. To capture such effects we set $f(\lambda) = 1 - (1-\lambda)^\alpha$, with $\alpha \geq 1$. As required $f(1) = 1$, and $f'(\lambda) = \alpha(1-\lambda)^{a-1} > 0$. We refer to $\alpha$ as the elasticity of employment loss with respect to the degree of social distancing, $\lambda$. The direct

15

employment effect of social distancing will be less adverse for values of $a > 1$. In what follows, in addition to the baseline value of $\alpha = 1$, we also consider $\alpha = 2$, under which a reduction of $\lambda$ from 1 to $1/2$, for example, reduces the employment rate by $1/4$ as compared to $1/2$ if we set $\alpha = 1$.

In addition to this direct effect, the employment level also falls directly due to the number of infected individuals, $I_t(\lambda)$, which also depends on $\lambda$. Recall that $I_t(\lambda)$ is increasing in $\lambda$. There will be more infected individuals the higher the level of exposure to the virus. Overall, the rate of employment during the spread of the epidemic is given by

$$e_t(\lambda) = f(\lambda)e_t - i_t(\lambda). \tag{20}$$

The associated employment loss is then

$$\ell_t(\lambda) = e_t - e_t(\lambda) = [1 - f(\lambda)] e_t + i_t(\lambda), \text{ for } t = 1, 2, ..., T. \tag{21}$$

This relationship represents a trade off between the opposing effects of high and low exposures to the epidemic on the rate of employment. In the event of a high exposure the first term of (21) will be small relative to the second term, and when exposure is low the direct employment loss is much higher than the indirect loss due to the spread of infection. It is also important to bear in mind that employment losses can vary considerably over the course of the epidemic.

Figures 4 shows the time profile of the simulated values of employment losses, $\ell_t(\lambda)$, for selected values of $\lambda = 0.5, 0.25$ and $0.1$, with $\alpha = 1$. We focus on exposure rates of 50% and less, since our estimates of $\lambda$ to be discussed tend to be rather small. The losses are computed with daily employment rates set to $e_t = 0.6$, which is approximately equal to the mean ratio of employment to population in the US during the last quarter of 2019. As before, the simulated values for $i_t(\lambda)$ are obtained using the SIR model with the parameters $R_0 = 3, \gamma = 1/d = 1/14$, and the initial values $i_1(\lambda) = \lambda/1000$, and $i_2(\lambda) = [1 - \gamma + (\gamma R_0/\lambda) s_1(\lambda)] i_1(\lambda)$, where $s_1(\lambda) = \lambda - i_1(\lambda)$.

As to be expected, employment losses mount up as the rate of exposure to the disease is reduced from 50% to 25% and right down to 10%. It is also evident from Figure 4 that with the flattening of the infection curve, as $\lambda$ is reduced, employment losses stabilize and remain high for the duration

of the epidemic. However, as noted earlier, the extent of the losses from reducing $\lambda$ very much depends on $\alpha$. As can be seen from Figure 4, when $a = 1$ the losses can be quite substantial, as everyone who is isolated has a full charge on the economy. When $a = 2$, the adverse effects of social distancing are somewhat mitigated, but could still be considerable for values of $\lambda$ below 25%. See Figure 5 which shows the simulated employment losses for $a = 1$ (blue) and $a = 2$ (red) with $\lambda = 0.25$.

**Figure 4:** Simulated employment losses for $\alpha = 1$ and the values of $\lambda = 0.5$ (blue), 0.25 (red) and 0.10 (green)
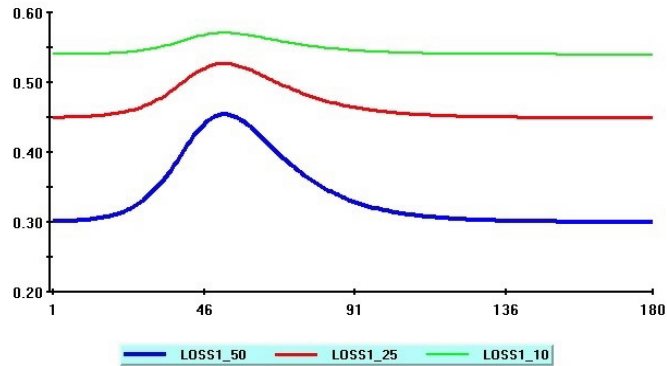


**Figure 5:** Simulated employment losses for $\lambda = 0.25$ and the values of employment loss elasticity, $a = 1$ (blue) and $a = 2$ (red)
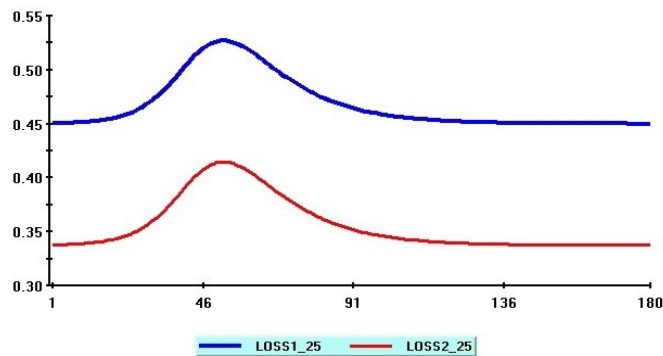
A summary of average simulated employment losses for different values of $\lambda$ and $\alpha$ is provided in Table 1. In this table we also give estimates for $\lambda = 1$ (case of no social distancing) and $\lambda = 0.75$ representing a moderate degree of social distancing. We also computed the simulated losses allowing for voluntary social distancing, but the results in Table 1 were not much affected. The results confirm that average losses over the duration of the epidemic (simulated to be 120 days) can be significant, but can be somewhat mitigated by working from home and by enabling a select group of workers to take part in productive activities under medical supervision (through regular testing for the disease) and by providing the necessary protective equipment for their own and other people's safety.

**Table 1**: Simulated average employment loss (in per cent per annum) due to epidemic under different social distancing ($\lambda$) and economic impact ($\alpha$) scenarios

| Employment loss elasticity | Social distancing coefficient ($\lambda$) | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|
| $\alpha$ | 1.0 | 0.75 | 0.50 | 0.25 | 0.10 |
| 1.0 | 3.6 | 7.7 | 11.8 | 15.9 | 18.4 |
| 1.5 | 3.6 | 5.2 | 8.9 | 13.9 | 17.4 |
| 2.0 | 3.6 | 4.0 | 6.8 | 12.2 | 16.6 |

Notes: This table reports results of a simulation of the epidemic under different social distancing ($\lambda$) and economic impact ($\alpha$) scenarios. The epidemic is simulated using SIR model with $R_0 = 3$ and $\gamma = 1/14$. $\lambda$ is the fraction of the population exposed to the virus. $\alpha$ determines the economic cost of the isolation measures, as defined by $(1 - \lambda)^a$. The losses are given in per cent per annum over 120 days which is the simulated length of the epidemic.

The calibration of $\lambda$ is a complicated undertaking and could differ across economies. In the case of the U.S., it is possible to estimate $\lambda$ from the recently stated aims by US administration to limit the number of fatalities due to the COIVD-19 to less than $200,000$. Assuming a death rate of $1\%$, this requires limiting the cumulative number of infected cases to $C^* = 200,000/0.01 = 20,000,000$. For a given $\lambda$ and the reproduction ratio of $R_0 = 3$, we have $C^*/P_{US} = 2\lambda/3$, which gives the estimate $\lambda_{US} = (20/320)(3/2) = 0.094$, assuming a US population of 320 million. The implied value of $\lambda_{US}$

would need to be even lower if a higher death rate is assumed, as the current US data suggests.[4] With $\lambda_{US}$ at 10%, the employment loss over the duration of the epidemic (estimated to be around 120 days), could be as much as 18.4% at an annual rate when $\alpha = 1$, but gets reduced to 16.6 per cent per annum when $\alpha = 2$.

# 5 Fitting the modified SIR model to the data: estimation of removal and exposure rates

Whilst calibration can be helpful in counterfactual analysis, it is also desirable to obtain estimates of $\lambda$ from realized outcomes. This is fortunately possible using data on COIVD-19 from Chinese provinces over the period January to March 2020. We also report estimates for a selected number of countries, but these estimates should be considered as preliminary since at the time of writing many of these epidemics are still unfolding. The attraction of using data from Chinese provinces is two-fold. First we have complete daily time series data that cover the full duration of the epidemics with slightly different start dates. Second, we can investigate the differences in parameter estimates (particularly $\lambda$) for the Hubei Province, the epicenter of the epidemic in China, as compared to the estimates for other provinces.

Our focus is on estimating the removal rate, $\gamma$, and the social distancing coefficient, $\lambda$. We base our estimation on equations (4) and the solution for $i_t$ given by (11). However, it is widely acknowledged that in the absence of large scale testing and given the asymptomatic nature of the disease in the case of many infected individuals, the recorded numbers of infected and recovered cases of COIVD-19 most definitely underestimate the true numbers of such cases. Before proceeding therefore we need to address this challenge.

## 5.1 Adjusting for under-reporting and other measurement errors

To allow for under-recording of infected cases, and other related measurement errors, we distinguish between the *true* and reported (confirmed) measures. We denote the true measures of infected and

---

[4] At the time of writing the death rate of COIVD-19 in the U.S. is around 4% using reported number confirmed cases. But as argued below, due to under reporting of infected cases the true death rate is likely to be around 2%.

recovered cases by $\tilde{C}_t$ and $\tilde{R}_t$, respectively, and denote the corresponding reported statistics by $C_t$ and $R_t$. Let $\pi_t$ be the ratio of confirmed to true cases, and suppose that

$$\pi_t = \pi e^{v_t - 0.5\sigma^2}, \tag{22}$$

where $\pi$ $(0 < \pi < 1)$ is a fixed fraction, and $v_t$ is $IIDN(0, \sigma^2)$. The assumption that $\pi_t$ follows a log-normal distribution is made for convenience and can be relaxed, and ensures that $E(\pi_t) = \pi$. It is also worth noting that $Var\left(\pi_t\right) = \pi^2(e^{\sigma^2} - 1)$. The inverse of $\pi$ measures of the degree of under-reporting and is referred to as the multiplication factor (MF) in the literature–see, for example, Gibbons et al. (2014). It is also reasonable to expect that the same fraction, $\pi_t$ applies to recovered cases. Under these assumptions we have

$$C_t = \pi_t \tilde{C}_t = \pi e^{v_t - 0.5\sigma^2} \tilde{C}_t, \quad \text{and} \quad R_t = \pi_t \tilde{R}_t = \pi e^{v_t - 0.5\sigma^2} \tilde{R}_t, \tag{23}$$

which also yield

$$I_t = C_t - R_t = \pi e^{v_t - 0.5\sigma^2} \tilde{I}_t, \tag{24}$$

where $I_t$ and $\tilde{I}_t$ are the reported and the true number of active cases, namely the number of individuals that remain infected in day $t$.

The theoretical equations (4) and (11) are derived in terms of the true measures, $\tilde{\imath}_t = \tilde{I}_t/P$, and $\tilde{r}_t = \tilde{R}_t/P$, but for estimation purposes they need to be cast in terms of the reported statistics, namely $i_t = I_t/P$ and $r_t = R_t/P$. Using (23) and (24) in the equation for the recovery rate, (4), we have

$$r_{t+1} = e^{v_{t+1} - v_t} \left(r_t + \gamma i_t\right),$$

which yields the following estimating equation

$$r_{t+1} = \rho \, r_t + (\gamma \rho) \, i_t + \varepsilon_{t+1}, \tag{25}$$

where $\rho = e^{\sigma^2}$, and under the assumption that $v_t$ are $IIDN(0, \sigma^2)$, it follows that $E\left(\varepsilon_{t+1} \left| r_t, i_t \right.\right) = 0$.

It is interesting to note that the MF, $1/\pi$, does not enter the equation for $r_{t+1}$, and both of the unknown parameters, $\sigma^2$ and $\gamma$ can be estimated from the OLS regression of $r_t$ on $r_{t-1}$ and $i_t$.

Similarly, using (24) to replace the true values $\tilde{i}_t$ in (11) we obtain

$$i_{t+1} = \left( i_t^2/i_{t-1} \right) \rho^3 + \left( \frac{\rho^2 \gamma R_0}{\pi \lambda} \right) \left[ i_t i_{t-1} \left( 1 - \gamma \right) - \rho i_t^2 \right] + \xi_{t+1}, \tag{26}$$

where $\rho = e^{\sigma^2}$, $E\left( \xi_{t+1} | i_t, i_{t-1} \right) = 0$. For a given value of $\rho$, $\gamma$ and $R_0$ the above non-linear regression can be used to provide estimates of $\pi \lambda$. Thus, to estimate $\lambda$ we need to make an assumption regarding, $\pi$, which we address below.

## 5.2 Estimates of recovery and social distancing rates

We allow recovery rates, $\gamma_j$, to differ across Chinese provinces reflecting possible differences in their demographics and the availability of medical facilities. Let $\gamma_j$ be the recovery rate in province $j$, and consider the regressions[5]

$$r_{j,t+1} = \rho_j \, r_{jt} + \left( \gamma_j \rho_j \right) i_{jt} + \varepsilon_{j,t+1}, \ \text{ for } j = 1, 2, ..., N, \tag{27}$$

where $r_{jt}$ and $i_{jt}$ are measured as

$$r_{jt} = (RE_{jt} + D_{jt})/P_j, \text{ and } I_{jt} = C_{jt} - RE_{jt} - D_{jt},$$

in which $C_{jt}, RE_{jt}$ and $D_{jt}$ are daily time series data obtained from Johns Hopkins University Carnivorous Resource Center, corresponding to the cumulative number of confirmed, recovered and deceased cases for province/country $j$, respectively.[6]

---

[5] Here the recovery rate includes both the recovered and the deceased, and strictly speaking should be referred to as the *removal rate*. But in line with the literature we use "recovery rate" in place of the "removal rate".

[6] Available at https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data. $C_{jt}$ is the number of confirmed cases taken from file time_series_covid19_confirmed_global.csv, $RE_{jt}$ is the number of recovered taken from file time_series_covid19_recovered_global.csv, and $D_{jt}$ is the number of deceased taken from file time_series_covid19_deaths_global.csv.

### 5.2.1 Estimates for Chinese provinces

The estimates of $\gamma_j$ for 21 Chinese provinces are summarized in Table 2.[7] As noted above, the estimation of $\gamma$ does not depend on under-recording of infected cases, but could depend on the random component of the measurement equations (23) and (24). The extent of this type of mis-measurement is given by the estimates of $\sigma_j$ which we also report in Table 2. Recall that $\sigma_j^2 = ln(\rho_j)$, and $\rho_j$, for $j = 1, 2, ..., 21$ are identified from regressions, (27). As can be seen, the estimates of recovery rates, $\gamma_j$, do not differ much across the provinces and lie in the range of 0.033 (for Beijing) and 0.066 (for Hebei), with the mean across all the provinces given by $\widehat{\gamma}_{MG} = 0.046$, with a standard error of 0.17%.[8] The random measurement errors are also relatively unimportant, with the estimates of $\sigma_j$ falling within the range $(0.06 - 0.10)$ across the 21 provinces.[9] The mean estimate, $\widehat{\gamma}_{MG} = 0.046$, corresponds to around 22 days on average from infection to recovery or death, with the rather narrow 95% confidence interval of 20 to 23 days. Hubei and Beijing have the lowest recovery rates, and Hebei, Hunan and Guinzhu provinces the highest. These estimates are all longer than the 14 days typically assumed in designing quarantine policies and, as we shall see, this has important implications for the estimates of the social distancing coefficient, $\lambda$.[10]

Next, we report estimates of $\lambda_j$ by running the non-linear regressions in (26) for each province separately. We consider two choices for $\rho$ and $\gamma$, namely the province-specific estimates, $\hat{\rho}_j$, $\hat{\gamma}_j$, reported in Table 2, and the pooled estimate, $\hat{\rho}_j = 1.0072$, $\widehat{\gamma}_{MG} = 0.046$. Regarding the choice of $R_0$, we consider 2.5 and 3, which are in the range of values reported in the recent report from Imperial College, Ferguson et al. (2020). But we only report the results for $R_0 = 3$, to save space. The estimates of $\lambda$ for other values of $R_0$ differ only in scale and can be easily obtained if desired.

---

[7]We dropped those provinces where the number of active cases did not exceed 100 during the period up to the end of February. This leaves 21 provinces: Hubei, Guangdong, Henan, Zhejiang, Hunan, Anhui, Jiangxi, Shandong, Jiangsu, Chongqing, Sichuan, Heilongjiang, Beijing, Shanghai, Hebei, Fujian, Guangxi, Shaanxi, Yunnan, Hainan, and Guizhou.

[8]The standard errors for the mean group estimator, $\widehat{\gamma}_{MG} = n^{-1}\sum_{j=1}^{n}\hat{\gamma}_j$, is computed using the formula given in Pesaran and Smith (1995), and as shown in Chudik and Pesaran (2019), they are robust to weak cross correlations across the provinces.

[9]It is indeed reassuring to note that all the 21 estimates of $\rho_j$, computed from separate regressions, are all larger than 1, and yield reasonable estimates for the standard error of the measurement errors, defined by $\sigma_j = \sqrt{ln(\rho_j)}$.

[10]The medical evidence documented in Ferguson et al. (2020) implies a value for $\gamma$ in the range 0.048 to 0.071, with our empirical evidence suggesting that values at the lower end of this range might be more appropriate.

Note from equation (26) that using time series data on $i_t$, we are only able to identify $\pi\lambda/R_0$, and other data sources must be used to set $R_0$ and $\pi$.

Finally, we calibrate the proportion of under reporting, $\pi$, or its inverse, $1/MF$, using the data from the Diamond Princess cruise ship reported in Morbidity and Mortality Weekly Report, Moriarty et al. (2020). This case could be viewed as quasi-experimental. It is reported that out of 3,711 passengers and crew 712 had positive test results for SARS-Cov-2, with only 381 being symptomatic with the remaining 311 asymptomatic at the time of testing. Since all of those on board are tested, it is reasonable to assume that the true number of infected individuals is $\tilde{C} = 712$, and in the absence of complete testing the confirmed number of symptomatic cases $C = 381$, since those without any symptoms would have been overlooked in the absence of complete testing. These statistics suggest $\hat{\pi} = C/\tilde{C} = 381/712 = 0.535$ or $MF = 1.9$, which we round to $MF = 2$. This estimate is preliminary but seems plausible when we consider the death rate reported for China and the death rate on Diamond Princess. It is widely recognized that the death rate of COIVD-19 based on confirmed infected cases could be grossly over-estimated, again due to under testing, and under estimation of $\tilde{C}$. It is therefore interesting to see if we can obtain a death rate for China which becomes closer to the death rate observed on Diamond Princess of 1.3%, if we adjust upward the number of confirmed cases by $MF = 2$. Based on confirmed cases and the number of deaths in China (at the time of writing) the crude death rate is $d_{China} = C/D = 3,335/81,865 = 4.07\%$. But assuming that $D = \tilde{D}$, and setting $MF = 2$, the true death rate in China reduces to

$$\tilde{d}_{China} = \frac{\tilde{D}}{\tilde{C}} = \frac{D}{C}\cdot\frac{C}{\tilde{C}} = 2.03\%.$$

This is still somewhat larger than the death rate of 1.3% reported for Diamond Princess, but could still be close to the truth, noting that on average access to medical facilities for the passengers and crews on Diamond Princess might be better as compared to China where most of the fatalities occurred at the epicenter of the epidemic in Hubei province without much warning or preparation at the start of the epidemic. In view of these results we set $\pi = 0.5$ and compute province-specific estimates of $\lambda_j$ with $R_0 = 3$ by running the regressions.[11]

---

[11] The death rate of COIVD-19 in other countries where the epidemic is still at its early stages is likely to be further

Estimates of $\lambda_j$ (per 100,000 of population) for Chinese provinces are reported in Table 3 using the full sample and a subsample. Initially, we focus on the full sample estimates. The left panel of Table 3 gives the results when the province-specific estimates of $\rho_j$ and $\gamma_j$ are used, whilst the right panel reports the estimates of $\lambda_j$ based on the pooled estimates of $\rho$ and $\gamma$, namely $\hat{\rho}_j = \hat{\rho}_{MG} = 1.0072$, and $\gamma_j = \hat{\gamma}_{MG} = 0.046$. The reason for including the pooled estimates is to investigate the robustness of the exposure rates, $\lambda_j$, to the choice of the recovery rate and the extent of measurement errors. As it turns out, both sets of estimates are quite close, with the ones based on the pooled estimates slightly larger. The largest difference between the two sets of estimates is obtained for Hubei province (the epicenter of the epidemic), namely 16.31 (1.76) when we use province-specific estimates of $\rho_j$ and $\gamma_j$, and 23.74 (2.74), when we use the pooled estimates ($\hat{\rho}_{MG}$ and $\hat{\gamma}_{MG}$).[12] Despite this difference, both estimates of $\lambda$ for Hubei province have a high degree of precision and are statistically highly significant, with their 95% confidence interval overlapping. What is striking is the large difference between the estimates for Hubei and the rest of the Chinese provinces. Outside the epicenter the estimates of $\lambda$ are much smaller in magnitude and range between 0.09 and 0.87, irrespective of whether we use province-specific or pooled estimates of $\rho$ and $\gamma$. In fact the mean group estimates of $\lambda$ across these provinces (excluding Hubei) are almost the same, namely 0.393 (0.05), and 0.389 (0.05), respectively. Thus on average the exposure rate in Hubei is estimated to be some $40 - 60$ times higher than the average exposure rate of the provinces outside of the epicenter. This makes sense, since it is likely that it took the Chinese authorities some time before they managed to put into effect very stringent social distancing polices that were required to reduce $\lambda$ substantially across China. Looking at the estimates of the province-specific exposure rates outside Hubei, we also see a remarkably low degree of heterogeneity consistent with a firm and homogenous imposition of social distancing policies, following what had been learned at the epicenter of the epidemic.

As can be seen from Figure 6, the regressions for $i_{jt}$ fit reasonably well and trace the epidemic curves accurately for *all* 21 Chinese provinces that we consider. Here, it is also notable that the time to the peak of the epidemic curve is about 4 weeks for most provinces, and the time to completion is

---

biased upward due to the long delay (4 weeks or more) between infection and death.

[12]The figures in brackets are standard errors of the estimates.

around 8 weeks. It is clear that epidemic curves this flat could not have materialized if it were not for the very stringent social distancing policies implemented by the Chinese authorities, as reflected in the very low estimates of $\lambda_j$ that we obtain, particularly once we consider provinces outside of the epicenter of the epidemic.

Thus far, we have focussed on Chinese data since they represent completed epidemic cycles across all 21 provinces. In contrast, at the time of writing, the peak of the epidemic has not been reached for other countries, where the first reported cases came a few weeks after China. So we now turn our attention to other countries, considering only those countries for which we have a sufficient history for a reliable estimation of $\lambda$. Before proceeding, for an accurate comparison with China, we report estimates of $\gamma$ and $\lambda$ for the same 22 Chinese provinces over a subsample ending on $20^{th}$ of February (covering the initial stages of the epidemic before reaching its peak). For this subsample the estimates of $\gamma$ (and $\rho$) are summarized in Table 4, which are directly comparable to the full-sample estimates in Table 2. As can be seen, for $\gamma$ we obtain much smaller estimates when we use the subsample as compared to the full sample, with a mean estimate of 0.018 compared to the full-sample estimate of 0.046, possibly reflecting the fact that before the peak of the epidemic the data do not capture the recoveries and deaths that will materialize in the following three weeks. It is therefore reasonable to expect that removal rates in other countries will converge to the very precise estimates for China that we have already reported for the full sample in Table 2.

Subsample estimates of $\lambda$ are presented in Table 3. For $\lambda$ the point estimates based on the two samples are very close, in line with the news reporting/anecdotal evidence of stringent and consistent implementation of social distancing. Remarkably, these estimates continue to be very precisely estimated. A visual comparison between the estimates of $\lambda$ based on the full and the subsamples is given in Figure 7. As can be seen, the estimates of $\lambda$ are essentially the same for the two sample periods, with the exception of the estimates for Shandong province where the subsample estimate of $\lambda$ is larger than the full-sample estimate. This could be due to the fact that outside the epicenter, Shandong is the only province to experience a second wave in mid course, as is evident from the plot of active cases for Shandong in Figure 6.

25

**Table 2**: Estimates of the recovery rates $(\gamma_j)$ for Chinese provinces

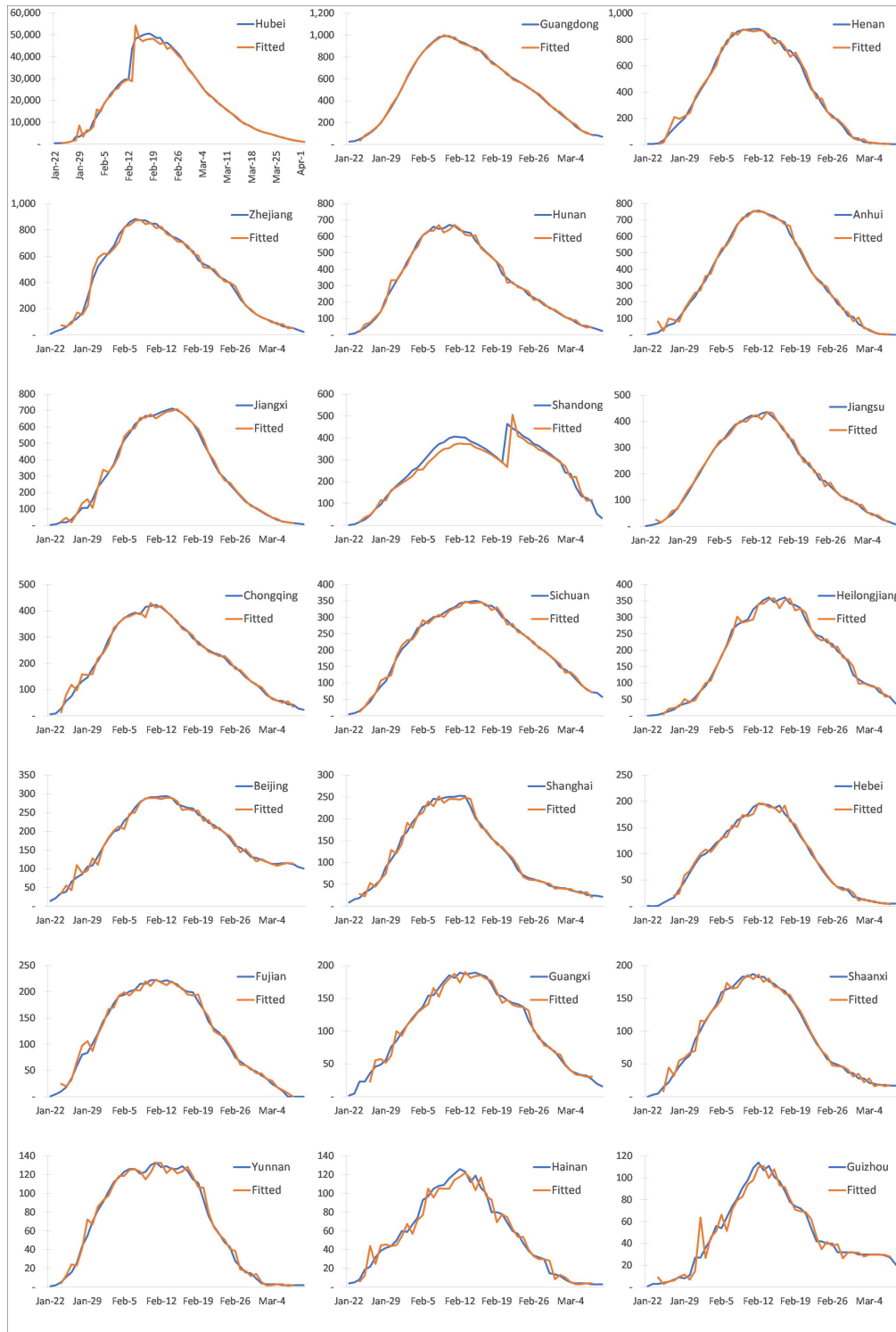| | $\hat{\gamma}_j$ | (s.e.) | $\hat{\rho}_j$ | (s.e.) | $\hat{\sigma}_j$ |
|---|---|---|---|---|---|
| Hubei | 0.035 | (0.0030) | 1.0088 | (0.0018) | 0.09 |
| Guangdong | 0.042 | (0.0032) | 1.0070 | (0.0018) | 0.08 |
| Henan | 0.054 | (0.0054) | 1.0072 | (0.0025) | 0.08 |
| Zhejiang | 0.044 | (0.0038) | 1.0073 | (0.0019) | 0.09 |
| Hunan | 0.056 | (0.0039) | 1.0055 | (0.0017) | 0.07 |
| Anhui | 0.049 | (0.0052) | 1.0081 | (0.0026) | 0.09 |
| Jiangxi | 0.047 | (0.0049) | 1.0081 | (0.0025) | 0.09 |
| Shandong | 0.046 | (0.0060) | 1.0084 | (0.0030) | 0.09 |
| Jiangsu | 0.055 | (0.0045) | 1.0063 | (0.0021) | 0.08 |
| Chongqing | 0.043 | (0.0039) | 1.0077 | (0.0021) | 0.09 |
| Sichuan | 0.040 | (0.0034) | 1.0081 | (0.0019) | 0.09 |
| Heilongjiang | 0.043 | (0.0052) | 1.0086 | (0.0030) | 0.09 |
| Beijing | 0.033 | (0.0042) | 1.0053 | (0.0028) | 0.07 |
| Shanghai | 0.043 | (0.0065) | 1.0034 | (0.0034) | 0.06 |
| Hebei | 0.066 | (0.0058) | 1.0060 | (0.0023) | 0.08 |
| Fujian | 0.039 | (0.0051) | 1.0078 | (0.0029) | 0.09 |
| Guangxi | 0.037 | (0.0051) | 1.0093 | (0.0031) | 0.10 |
| Shaanxi | 0.044 | (0.0051) | 1.0074 | (0.0027) | 0.09 |
| Yunnan | 0.041 | (0.0068) | 1.0085 | (0.0038) | 0.09 |
| Hainan | 0.052 | (0.0078) | 1.0080 | (0.0037) | 0.09 |
| Guizhou | 0.056 | (0.0079) | 1.0049 | (0.0038) | 0.07 |
| MG estimates | 0.046 | (0.0017) | 1.0072 | (0.0003) | 0.08 |

Notes: Estimation is based on regression $r_{j,t+1} = \rho_j\ r_{jt} + \left(\gamma_j\rho_j\right)i_{jt} + \varepsilon_{j,t+1}$, where $\sigma_j = \sqrt{ln(\rho_j)}$. See also (27). Sample is Jan-22 to March-31, 2020 ($T = 70$) with the exception of Hubei which is estimated using the sample Jan-22 to Apr-6 ($T = 76$).

**Table 3**: Estimates of exposure rates ($\lambda_j$) per 100,000 population for Chinese provinces

$(\pi = 0.5,\ R_0 = 3)$

| | Using province-specific estimates $\hat{\rho}_j, \hat{\gamma}_j$ | | | | Using pooled estimates $\hat{\rho}_{MG}, \hat{\gamma}_{MG}$ | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Full sample | | Subsample | | Full sample | | Subsample | |
| | $\widehat{\lambda_j}$ | (s.e.) | $\widehat{\lambda_j}$ | (s.e.) | $\widehat{\lambda_j}$ | (s.e.) | $\widehat{\lambda_j}$ | (s.e.) |
| Guangdong | 0.441 | (0.042) | 0.431 | (0.044) | 0.491 | (0.044) | 0.484 | (0.046) |
| Henan | 0.610 | (0.099) | 0.571 | (0.115) | 0.501 | (0.085) | 0.457 | (0.093) |
| Zhejiang | 0.610 | (0.079) | 0.599 | (0.110) | 0.643 | (0.084) | 0.633 | (0.117) |
| Hunan | 0.663 | (0.108) | 0.677 | (0.146) | 0.468 | (0.072) | 0.472 | (0.095) |
| Anhui | 0.734 | (0.159) | 0.692 | (0.157) | 0.722 | (0.168) | 0.675 | (0.164) |
| Jiangxi | 0.846 | (0.166) | 0.770 | (0.176) | 0.868 | (0.180) | 0.786 | (0.190) |
| Shandong | 0.087 | (0.008) | 0.200 | (0.034) | 0.088 | (0.008) | 0.216 | (0.040) |
| Jiangsu | 0.389 | (0.065) | 0.399 | (0.075) | 0.297 | (0.049) | 0.300 | (0.057) |
| Chongqing | 0.586 | (0.106) | 0.583 | (0.133) | 0.644 | (0.118) | 0.642 | (0.150) |
| Sichuan | 0.166 | (0.021) | 0.165 | (0.026) | 0.211 | (0.027) | 0.213 | (0.035) |
| Heilongjiang | 0.364 | (0.050) | 0.364 | (0.057) | 0.414 | (0.060) | 0.420 | (0.072) |
| Beijing | 0.412 | (0.077) | 0.444 | (0.127) | 0.586 | (0.096) | 0.644 | (0.163) |
| Shanghai | 0.498 | (0.106) | 0.524 | (0.156) | 0.451 | (0.079) | 0.472 | (0.115) |
| Hebei | 0.215 | (0.050) | 0.205 | (0.070) | 0.126 | (0.029) | 0.115 | (0.037) |
| Fujian | 0.211 | (0.042) | 0.194 | (0.041) | 0.266 | (0.053) | 0.251 | (0.054) |
| Guangxi | 0.106 | (0.030) | 0.104 | (0.045) | 0.156 | (0.049) | 0.156 | (0.076) |
| Shaanxi | 0.213 | (0.040) | 0.199 | (0.050) | 0.227 | (0.043) | 0.213 | (0.054) |
| Yunnan | 0.108 | (0.020) | 0.095 | (0.019) | 0.130 | (0.026) | 0.116 | (0.025) |
| Hainan | 0.461 | (0.079) | 0.463 | (0.108) | 0.396 | (0.069) | 0.395 | (0.094) |
| Guizhou | 0.131 | (0.034) | 0.133 | (0.046) | 0.096 | (0.023) | 0.096 | (0.031) |
| MG estimate | 0.393 | (0.050) | 0.391 | (0.047) | 0.389 | (0.050) | 0.388 | (0.047) |
| Hubei | 16.306 | (1.764) | 16.164 | (2.765) | 23.743 | (2.720) | 23.584 | (4.395) |

Notes: Estimation is based on regressions $i_{j,t+1} = \left( i_{jt}^2/i_{j,t-1} \right) \rho_j^3 + \rho_j^2 \left( \frac{\gamma_j R_0}{\pi \lambda_j} \right) \left[ i_{jt} i_{j,t-1} \left( 1 - \gamma_j \right) - \rho_j i_{jt}^2 \right] + \xi_{jt+1}$, for $j = 1, 2, ..., 21$ (provinces), with $\rho_j, \gamma_j$ imposed equal to the country-specific estimates from Table 2 or their MG estimates from Table 2. The full sample is the same as in Table 2. The subsample is Jan-22 to Feb-20 ($T = 30$).

**Figure 6**: Number of active cases in Chinese provinces ($I_{jt}$) and fitted values

**Table 4**: Subsample estimates of recovery rates ($\gamma_j$) for Chinese provinces using the sample ending Feb-20

| | $\hat{\gamma}_j$ | (s.e.) | $\hat{\rho}_j$ | (s.e.) | $\hat{\sigma}_j$ |
|---:|---|---|---|---|---|
| Hubei | 0.013 | (0.0038) | 1.0656 | (0.0226) | 0.25 |
| Guangdong | 0.022 | (0.0037) | 1.0587 | (0.0104) | 0.24 |
| Henan | 0.025 | (0.0052) | 1.0636 | (0.0135) | 0.25 |
| Zhejiang | 0.024 | (0.0047) | 1.0492 | (0.0115) | 0.22 |
| Hunan | 0.031 | (0.0064) | 1.0581 | (0.0124) | 0.24 |
| Anhui | 0.009 | (0.0042) | 1.1391 | (0.0161) | 0.36 |
| Jiangxi | 0.006 | (0.0037) | 1.1422 | (0.0150) | 0.36 |
| Shandong | 0.019 | (0.0045) | 1.0739 | (0.0140) | 0.27 |
| Jiangsu | 0.019 | (0.0049) | 1.1020 | (0.0129) | 0.31 |
| Chongqing | 0.019 | (0.0038) | 1.0840 | (0.0109) | 0.28 |
| Sichuan | 0.013 | (0.0062) | 1.0823 | (0.0208) | 0.28 |
| Heilongjiang | 0.016 | (0.0077) | 1.0785 | (0.0373) | 0.27 |
| Beijing | 0.014 | (0.0063) | 1.0681 | (0.0231) | 0.26 |
| Shanghai | 0.020 | (0.0089) | 1.0819 | (0.0226) | 0.28 |
| Hebei | 0.026 | (0.0073) | 1.0830 | (0.0164) | 0.28 |
| Fujian | 0.012 | (0.0050) | 1.0916 | (0.0199) | 0.30 |
| Guangxi | 0.009 | (0.0079) | 1.1034 | (0.0373) | 0.31 |
| Shaanxi | 0.010 | (0.0042) | 1.1194 | (0.0156) | 0.34 |
| Yunnan | 0.002 | (0.0080) | 1.1609 | (0.0382) | 0.39 |
| Hainan | 0.027 | (0.0144) | 1.0621 | (0.0373) | 0.25 |
| Guizhou | 0.038 | (0.0140) | 1.0434 | (0.0349) | 0.21 |
| MG estimate | 0.018 | (0.0019) | 1.0862 | (0.0067) | 0.29 |

Notes: Estimation is based on regression $r_{j,t+1} = \rho_j \, r_{jt} + \left(\gamma_j \rho_j\right) i_{jt} + \varepsilon_{j,t+1}$. See also (27). Sample is Jan-22 to Feb-20 ($T = 30$).

**Figure 7**: Full sample and subsample estimates of exposure rate ($\lambda$) per 100,000 population across Chinese provinces ($\pi = 0.5$, $R_0 = 3$)
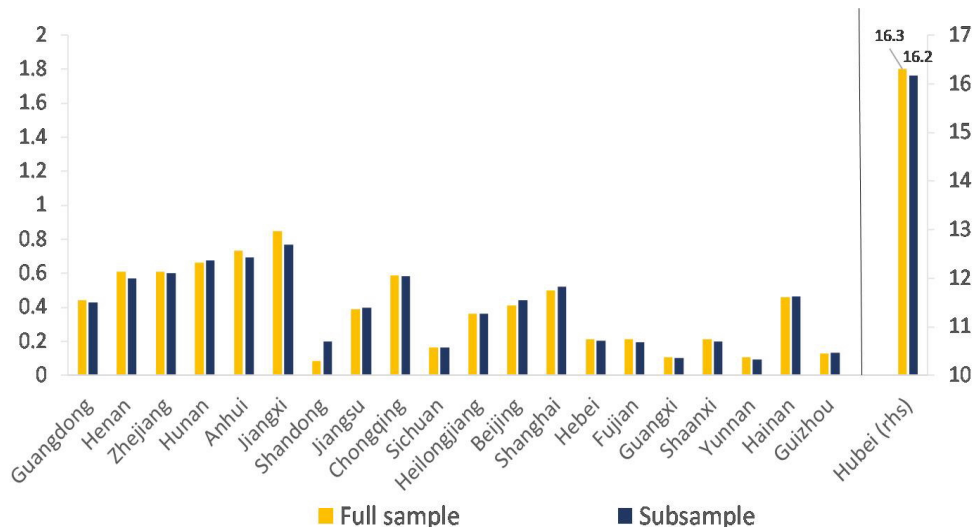


Note: Full sample results are based on the same samples as in Table 2. The subsample estimates are based on the sample Jan-22 to Feb-20, 2020. $\rho_j, \gamma_j$ are country-specific as estimated in Table 2.

## 5.3 Estimation results for selected countries

Bearing in mind the preliminary nature of these estimates using a subsample that covers the first phase of the epidemic, we now provide estimates for a selected number of countries with at least 30 days of reported active cases. Specifically, we consider South Korea, Iran and a selected number of European countries (Spain, Italy, France, Germany, United Kingdom, Belgium, Switzerland, and Austria).[13] Additional countries can be included as the epidemic spreads and moves to its later stages in the rest of the world.

The estimates of $\gamma$ and $\rho$ for these countries are summarized in Table 5. Estimates of $\gamma$ are generally lower than the ones we obtain for Chinese provinces, using the full sample, implying longer removal times. In some cases, they are also less precisely estimated. Nevertheless, these country estimates are remarkably close to the estimates we obtained in Table 4 for Chinese provinces, when

---

[13] The fitting of epidemic curves for the United States or its counties is hampered by the lack of data for recovered cases. Due to the absence of minimally reliable data on recoveries, JHU stopped reporting these statistics at the US state level on March 9, 2020. Given the large variation in social distancing policies across US states and counties, it would be misleading to report a country-wide estimate of $\lambda$ for the country as a whole.

we used the subsample. It is therefore reasonable to conjecture that the mean estimate of $\gamma$ across European countries and elsewhere will end up converging to the mean estimate of $\gamma = 0.046$ that we have provided for China using the full samples.

The estimates of exposure rates, $\lambda$, for the selected countries are summarized in Table 6, with the left panel giving the estimates using country-specific estimates of $\rho$ and $\gamma$ from Table 5, and the right panel using the pooled estimate of $\rho$ and $\gamma$ from Chinese provinces, namely $\hat{\rho}_{MG} = 1.0072$, $\hat{\gamma}_{MG} = 0.046$, which are arguably closer to their true values. This choice is driven by our belief that the pooled estimates based on the full set of data from Chinese provinces provide a more reliable benchmark for these parameter values for countries with advanced scientific and medical capabilities and are possibly less likely to be biased as compared to the MG estimate for the selected countries in Table 5.

As can be seen, there is a great deal of heterogeneity in the estimates of $\lambda$ across the selected countries, partly reflecting differences in the mitigation polices adopted. However, most of these estimates are much larger than those obtained for Chinese provinces ex-Hubei. Strikingly, and unfortunately, these estimates are 3-6 times larger than in Hubei, reflecting a massive failure at learning from the Chinese earlier experience with firm and uniform social distancing. With the exception of Iran, South Korea, and Austria the estimates of $\lambda$ for the European countries in our sample are all either comparable or much higher than estimates we have obtained for the Hubei province. In particular, Italy, Spain and Belgium stand out with the largest estimates, all more than multiple times as high as Hubei. Again, remarkably, even though less precisely estimated than in the case of Chinese provinces sub-samples, these estimates continue to be relatively precise. We therefore have confidence that, as more data will become available, a very precise estimate of the actual degree of social distancing adopted will be measurable from the data.

**Table 5** Estimated values of recovery rates $(\gamma_j)$ for a selection of countries

|  | $\hat{\gamma}_j$ | (s.e.) | $\hat{\rho}_j$ | (s.e.) | $\hat{\sigma}_i$ | Sample | $T$ |
|---|---|---|---|---|---|---|---|
| Spain | 0.054 | (0.0015) | 1.000 | . | . | 25-Feb to 12-Apr | 46 |
| Italy | 0.026 | (0.0010) | 1.000 | . | . | 20-Feb to 12-Apr | 51 |
| France | 0.035 | (0.0025) | 1.000 | . | . | 26-Feb to 12-Apr | 45 |
| Germany | 0.045 | (0.0124) | 1.023 | (0.022) | 0.15 | 24-Feb to 12-Apr | 47 |
| United Kingdom | 0.016 | (0.0006) | 1.000 | . | . | 24-Feb to 12-Apr | 47 |
| Iran | 0.007 | (0.0171) | 1.081 | (0.020) | 0.28 | 20-Feb to 12-Apr | 51 |
| Belgium | 0.041 | (0.0056) | 1.003 | (0.013) | 0.05 | 02-Mar to 12-Apr | 40 |
| Switzerland | 0.037 | (0.0098) | 1.039 | (0.018) | 0.20 | 28-Feb to 12-Apr | 43 |
| Austria | 0.023 | (0.0037) | 1.077 | (0.010) | 0.27 | 28-Feb to 12-Apr | 43 |
| South Korea | 0.022 | (0.0071) | 1.020 | (0.010) | 0.14 | 19-Feb to 12-Apr | 52 |
| MG estimate | 0.031 | (0.0043) | | | | | |

Notes: Estimation is based on regression $r_{j,t+1} = \rho_j\, r_{jt} + (\gamma_j\rho_j)\, i_{jt} + \varepsilon_{j,t+1}$.

**Table 6**: Estimated values of exposure rate $(\lambda_j)$ per 100,000 population for a selection of countries $(\pi = 0.5,\ R_0 = 3)$

|  | Using country-specific estimates $\hat{\rho}_j, \hat{\gamma}_j$ | | Using pooled estimates $\hat{\rho}_{MG}, \hat{\gamma}_{MG}$ | |
|---|---|---|---|---|
|  | $\widehat{\lambda}_j$ | (s.e.) | $\widehat{\lambda}_j$ | (s.e.) |
| Spain | 457.9 | (256.9) | 117.8 | (19.5) |
| Italy | 252.4 | (153.8) | 117.8 | (12.7) |
| France | 25.8 | (4.6) | 34.5 | (5.5) |
| Germany | 23.6 | (1.7) | 43.9 | (6.6) |
| United Kingdom | 47.2 | (18.2) | 79.6 | (13.0) |
| Iran | 0.8 | (0.1) | 17.3 | (2.9) |
| Belgium | 172.9 | (52.7) | 120.8 | (20.9) |
| Switzerland | 24.8 | (1.5) | 65.4 | (10.9) |
| Austria | 6.8 | (0.3) | 47.7 | (6.9) |
| South Korea | 1.6 | (0.2) | 5.5 | (0.8) |
| MG estimate | 101.4 | (37.4) | 65.0 | (23.7) |

Notes: Estimation is based on regression $i_{j,t+1} = \left(i_{jt}^2/i_{j,t-1}\right)\rho_j^3 + \rho_j^2\left(\frac{\gamma R_0}{\pi\lambda}\right)\left[i_{jt}i_{j,t-1}\left(1-\gamma\right) - \rho i_t^2\right] + \xi_{t+1}$, using the same sample as in Table 3. The left part of this table reports results using country-specific estimates $\hat{\rho}_j, \hat{\gamma}_j$, and the right part of this table uses pooled estimates $\hat{\rho}_{MG}, \hat{\gamma}_{MG}$ from Table 2

# 6    Conclusions

This paper makes two related contributions. At the theoretical level, it integrates social distancing polices in a standard SIR model in order to evaluate their impact on both the COVID-19 epidemic and the associated employment costs. The framework distinguishes between mandatory and voluntary isolation. Critically, it is shown with simulations that, while targeted mandated policies can be very useful in flattening the epidemic curve, voluntary policies are relatively ineffective. Self-isolation can affect the epidemic curve, but only when it is close to its peak. We also show that mandating social distancing is very effective at flattening the epidemic curve, but is costly in terms of employment loss. However, if targeted towards individuals most likely to spread the infection, the employment loss can be somewhat reduced.

At the empirical level, using JHU daily COVID-19 statistics corrected for measurement errors, the paper provides estimates of province-specific recovery ($\gamma$) and exposure ($\lambda$) rates in China and in a selected number of countries, and shows that the rate of exposure to COVID-19 was around $40 - 60$ times higher in Hubei at the epicenter of the epidemic compared to the rest of China. We find a very high degree of effective isolation, stable over time, and homogeneous across Chinese provinces. In contrast, we document lower and more heterogeneous degrees of effective isolation across European countries, and the degree of effective isolation turns out to be the lowest in Italy and Spain with an exposure some five times larger than our estimate for Hubei province, the epicenter of the epidemic in China.

As more and more reliable data becomes available, extending the empirical analysis to the United States and its metropolitan areas, as well as other countries worldwide, is an essential area of further research.

A further challenge is to relate the elasticity of employment loss, $\alpha$, to the way social distancing policies are implemented, including intensive testing and contact tracing. These are important topics of current and future research.

# References

Atkeson, A. (2020). What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios. NBER Working Paper No. 26867, https://doi.org/10.3386/w26867.

Barro, R. J., J. F. Ursua, and J. Weng (2020). The coronavirus and the great influenza pandemic: Lessons from the "Spanish flu" for the coronavirus's potential effects on mortality and economic activity. NBER Working Paper No. 26866, https://doi.org/10.3386/w26866.

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2020). An SEIR infectious disease model with testing and conditional quarantine. NBER Working Paper No. 26901, https://doi.org/10.3386/w26901.

Chudik, A. and M. H. Pesaran (2019). Mean group estimation in presence of weakly cross-correlated estimators. *Economics Letters 175*, 101 – 105. https://doi.org/10.1016/j.econlet.2018.12.036.

Correia, S., S. Luck, and E. Verner (2020). Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu. Available at SSRN: https://ssrn.com/abstract=3561560 or https://doi.org/10.2139/ssrn.3561560.

Diekmann, O. and J. Heesterbeek (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Son, New York, ISBN: 0471986828.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2020). The macroeconomics of epidemics. NBER Working Paper No. 26882, https://doi.org/10.3386/w26882.

Fang, H., L. Wang, and Y. Yang (2020). Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China. NBER Working Paper No. 26906, https://doi.org/10.3386/w26906.

Ferguson, N., P. Walker, C. Whittaker, and et al (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London COVID-19 Reports, Report No. 9, 16 March 2020, available at https://doi.org/10.25561/77482.

Flaxman, S., S. Mishra, A. Gandy, and et al. (2020). Estimating the number of infections and the impact of nonpharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London COVID-19 Reports, Report No. 13, 30 March 2020, https://doi.org/10.25561/77731.

Gibbons, C. L., M.-J. J. Mangen, D. Plass, and et al (2014). Measuring underreporting and underascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health 14:147*, 1471–2458. https://doi.org/10.1186/1471-2458-14-147.

Harko, T., F. S. Lobo, and M. Mak (2014). Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation 236*, 184–194. https://doi.org/10.1016/j.amc.2014.03.030.

Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London 22*(772), 700–721. https://doi.org/10.1098/rspa.1927.0118.

Kucharski, A. J., T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R. M. Eggo (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(20)30144-4.

Li, Q., X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, ..., and Z. Feng (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine 382*(13), 1199–1207. https://doi.org/10.1056/NEJMoa2001316.

Linton, O. (2020). When will the Covid-19 pandemic peak? Cambridge Working Papers in Economics CWPE2025, available at: http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe2025.pdf.

Metz, J. A. J. (1978). The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheoretica 27*, 75–123. https://doi.org/10.1007/bf00048405.

Moriarty, L., M. Plucinski, B. Marston, and et al. (2020). Public health responses to COVID-19 outbreaks on cruise ships - worldwide, February-March 2020. Morbidity and Mortality Weekly Report (MMWR), 26 March 2020, 69:347-352. https://doi.org/10.15585/mmwr.mm6912e3.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*(1), 79–113. https://doi.org/10.1016/0304-4076(94)01644-F.

Riou, J. and C. L. Althaus (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-ncov), December 2019 to January 2020. *Eurosurveillance 25*(4). https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058.

Salje, H., D. A. Cummings, and J. Lessler (2016). Estimating infectious disease transmission distances using the overall distribution of cases. *Epidemics 17*, 10–18. https://doi.org/10.1016/j.epidem.2016.10.001.

Satsuma, J., R. Willox, A. Ramani1, B. Grammaticos, and A. Carstea (2004). Extending the SIR epidemic model. *Physica A: Statistical Mechanics and its Applications 336*, 369–375. https://doi.org/10.1016/j.physa.2003.12.035.

Stock, J. H. (2020). Random testing is urgently needed. Manuscript, available at: http://www.igmchicago.org/covid-19/random-testing-is-urgently-needed/.

Wang, C., L. Liu, X. Hao, H. Guo, Q. Wang, J. Huang, ..., and T. Wu (2020). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. *medRxiv*. https://doi.org/10.1101/2020.03.03.20030593.