NBER WORKING PAPER SERIES

AGGREGATE AND FIRM-LEVEL STOCK RETURNS DURING PANDEMICS, IN REAL TIME

Laura Alfaro
Anusha Chari
Andrew N. Greenland
Peter K. Schott

Working Paper 26950
http://www.nber.org/papers/w26950

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

This paper is preliminary and incomplete. Missing citations and discussions of related research will be added in future drafts. We thank Nick Barberis, Lorenzo Caliendo, Teresa Fort, Mihai Ion, and Ed Kaplan for comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Aggregate and Firm-Level Stock Returns During Pandemics, in Real Time
Laura Alfaro, Anusha Chari, Andrew N. Greenland, and Peter K. Schott
NBER Working Paper No. 26950
April 2020
JEL No. E27,F1,G12

ABSTRACT

We show that unanticipated changes in predicted infections during the SARS and COVID-19 pandemics forecast aggregate equity market returns. We model cumulative infections as either exponential or logistic, and re-estimate the parameters of these models each day of the outbreak using information reported up to that day. For each trading day t we compute the change in predicted infections using day t – 1 versus day t – 2 information. Regression results imply that a doubling of such predictions is associated with a 4 to 11 percent decline in aggregate market value. This result implies a decline in returns' volatility as the trajectory of the pandemic becomes clearer.

Laura Alfaro
Harvard Business School
Morgan Hall 263
Soldiers Field
Boston, MA 02163
and NBER
lalfaro@hbs.edu

Anusha Chari
301 Gardner Hall
CB#3305, Department of Economics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
and NBER
achari@unc.edu

Andrew N. Greenland
Elon University
50 Campus Drive
Elon, NC 27244
agreenland@elon.edu

Peter K. Schott
Yale School of Management
165 Whitney Avenue
New Haven, CT 06511
and NBER
peter.schott@yale.edu

1 Introduction

Pandemics inflict a substantial toll. They also roil asset markets. In this paper, we show that unanticipated changes in predicted infections based on daily re-estimation of simple models of infectious disease forecast stock returns. This relationship is consistent with investors using such models to update their beliefs about the economic severity of the outbreak, in real time, as they attempt to gauge risk in the face of substantial uncertainty (Knight, 1921; Keynes, 1937).

We emphasize that we are *not* epidemiologists and are *not* outlining a method to characterize the true path of pandemics, or infer the efficacy of various intervention strategies.[1] Such efforts, while of immense value, require data which may not be available until after the outbreak is substantially underway. Rather, we view real-time changes in the predicted severity of an outbreak as potentially useful summary statistics of its consequences, especially before the true model is revealed.

We model cumulative infections as either exponential or logistic. We re-estimate the parameters of these models each day of the outbreak using information on the trajectory of reported cases up to that day, which arrives after trading closes on that day. We then use these parameters to compute the predicted number of cases for trading day t using the cumulative counts reported after closing on days $t-1$ and $t-2$. The difference in these forecasts for day t, therefore, reflect changes in expectations about the trajectory of the pandemic based on the newly available information. We then examine how these differences in projections covary with aggregate market returns on day t.

Applying this procedure to the 2003 SARS outbreak in Hong Kong and the current United States COVID-19 pandemic, we find that sharper increases in predictions are associated negatively with larger swings in market returns. Coefficient estimates imply declines of 8 to 11 percent in the Hang Seng index in Hong Kong in response to a doubling of predicted cases during the SARS outbreak. For the United States during COVID-19, we find, thus far, a similar relationship: a doubling of predicted cases implies an aggregate market decline of 4 to 10 percent in the Wilshire 5000 index. These findings suggest equity markets are less responsive to new cases the more they adhere to *previously* estimated parameters.

We find that changes in forecasts retain their explanatory power even after controlling for a simpler summary of the severity of the outbreak, the most recent increase in reported cases. In contrast to this simple measure, estimated model parameters explicitly predict the eventual number of people that may be infected (e.g., the "carrying capacity" under the logistic model), and the speed with which that number may be reached. For example, a jump in estimated share of the population that ultimately will be infected suggests a larger labor supply shock, while an increase in the estimated growth rate of infections has implications for healthcare capacity constraints.

While our results at present focus on SARS and COVID-19 in Hong Kong and the United States, we are expanding the set of countries under study, and have begun a similar analysis for the 2009 H1N1 outbreaks. We also have begun investigating the link between returns and exposure to pandemics at the firm level. Such exposure may vary along various channels, including domestic and international input-output linkages, as well as the demographics and occupations of firms' labor forces.

Our analysis contributes to several literatures. First, we add to the very large body of research on asset pricing that examines the predictability of stock returns. Seminal papers by Campbell and Shiller (1988), Fama and French (1988) and others show that factors ranging from valuation ratios to corporate payout and financing policies forecast stock returns. In this paper we draw upon standard epidemiological models to infer how investors might update their beliefs about disease progression.

[1] Piguillem and Shi (2020) and Berger et al. (2020), by contrast, use estimates of micro-founded SIR models to argue that expanded testing generates substantial welfare gains relative to quarantines.

Second, our planned examination of firm returns in response to changes in model predictions contributes to numerous studies in corporate finance, pioneered by Ball and Brown (1968) and Fama et al. (1969), which use plausible changes in investors' information sets to understand market dynamics. In a typical event study, researchers examine specific events, such as an earnings announcements, that may release information relevant to investors' beliefs about firm market value. Firms' "abnormal" returns relative to a benchmark asset pricing model during such events summarize these changes expectations.[2] Here, we demonstrate that plausibly exogenous changes in the daily information set regarding the epidemic's trajectory are correlated with firms' stock returns.[3]

Third, our paper contributes to the very large literature in public health which attempts to explain the trajectory of infections during a pandemic.[4] In contrast to that research, we link changes in the estimated parameters and predictions of these models in real time to economic outcomes. To reiterate, we do not claim that the evolution of a pandemic must follow a purely exponential or logistic growth path. Rather, we explore whether the predictions of these models are informative of economic conditions, as manifest in their correlation with the market.[5] An interesting question for further research is the extent to which feedback from the predicted health and economic consequences of the outbreak affects future infections. For example, dire enough anticipated economic consequences might influence the set of policies used to combat the outbreak, thereby altering its trajectory (Lucas, 1976).

Finally, this paper relates to a rapidly emerging literature studying the economic consequences of COVID-19, and a more established literature investigating earlier pandemics. Barro et al. (2020), for example, argue that the decline in output during the 1918 to 1920 "Spanish Flu" epidemic provide a plausible mode of the economic consequences of COVID-19. Our analysis complements (Ramelli and Wagner, 2020), who focus on debt and international supply chains as key channels of exposure to the COVID-19 epidemic, and Gormsen and Koijen (2020), who use the performance of US futures' markets during the outbreak to infer bounds on future GDP growth.

This paper proceeds as follows. Section 2 provides a brief description of infectious disease models and how investors might link the predictions of these models and to asset prices. Section 4 applies our framework to COVID-19. Section 5 concludes.

2 Modeling

In this section we outline how infectious disease outbreaks can be modeled in real time, and how investors might make use of the model's estimated parameters.

2.1 Epidemiological Models of Infectious Diseases

Exponential and logistic growth models are frequently used in biology and epidemiology to model infection and mortality. An exponential model,

$$C_{it} = a_i e^{(r_i t)} \tag{1}$$

[2] Wang et al. (2013), for example, examines how the stocks of Taiwanese biotechnology companies respond to a series of infectious disease outbreaks.

[3] Greenland et al. (2019) exploit a change in US trade policy to identify firms' exposure to greater import competition from China.

[4] Early contributions to this literature include Ross (1911), Kermack and McKendrick (1927), Kermack and McKendrick (1937) and Richards (1959).

[5] For an interesting discussion on the complexities associated with modeling an outbreak in real time, see https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-covid-19-model/.

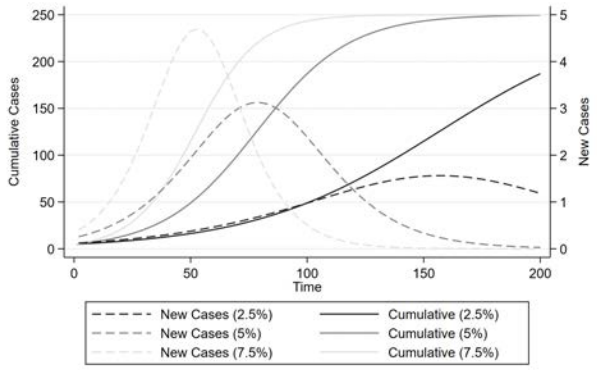
predicts the cumulative number of cases in country i on day t, C_{it}, as a function of the growth rate of infections in that country, r_i, the initial number of infected persons a_i, and time. In an exponential model, the number of infections per day continues to climb indefinitely. While clearly unrealistic ex-post, the exponential growth model is consistent with early stage pandemic growth rates.

In a logistic model (Richards, 1959), by contrast, the growth in infections grows exponentially initially, but then declines as the stock of infections approaches the population's "carrying-capacity," i.e., the cumulative number of people that ultimately will be infected. Carrying capacity is generally less than the full population. In a logistic model, the cumulative number of infections for country i on day t is given by:

$$C_{it} = \frac{k_i}{1 + c_i e^{(-r_i t)}}, \tag{2}$$

where k_i is the carrying capacity for country i, c_i is a shift parameter (characterizing the number of initially infected persons in country i) and r_i is the growth rate. Figure 1 provides an example of logistic infections for three different growth rates (2.5%, 5% and 7.5%) assuming $k_i = 250$ and $c_i = 50$. For each growth rate, we plot both the cumulative number of cases as of each day (left axis) and the number of new cases each day (right axis). As indicated in the figure, higher growth rates both shorten the time required to reach carrying capacity, and increase the peak number of infections.

Figure 1: Example of Logistic Pandemic with Different Rates of Infection



Source: Authors' calculations. Figure compares new and cumulative infections from days 1 to 200 assuming a logistic model with $k_i = 250$ and $c_i = 50$ and noted growth rates (r_i).

Given data on the actual evolution of infections, the two parameters in equation 1 and the three parameters in equation 2 can be updated each day using the sequence of infections up to that date. We estimate these sequences using STATA's nonlinear least squares command (`nl`).[6] STATA's `nl` command requires a vector of starting values, one each for each parameter to be estimated.

We encounter two problems during our estimation of logistic functions in our applications below. First, estimates for each day t are sensitive to the choice of starting values for that day, particularly in the initial days of the pandemic. This feature of the estimation is not surprising: when the number of cases is relatively small, the data are consistent with a wide range of logistic curves, and the objective function across them may be relatively flat.

[6] We are exploring other estimation procedures for use in a future draft, including use of SIR and SEIR models, e.g., Li et al. (2020) and Atkeson (2020).

To increase the likelihood that our parameter estimates represent the *global* solution, we estimate 500 epidemiological models for each day, 250 for the logistic case, and 250 for the exponential case. In each iteration we use a different vector of starting values. For each day t, our first starting values are the estimated coefficients from the prior day, if available.[7] In the case of the logistic model, we then conduct a grid search defined by all triples $\{r, c, k\}$ such that

$$r \in \{0.01, 0.21, 0.41, 0.61, 0.81\}$$
$$c \in \{\widehat{c_i^{t-1}}, 2 * \widehat{c_i^{t-1}}, 4 * \widehat{c_i^{t-1}}, \ldots, 10 * \widehat{c_i^{t-1}}\}$$
$$k \in \{\widehat{k_i^{t-1}}, 2 * \widehat{k_i^{t-1}}, 3 * \widehat{k_i^{t-1}}, \ldots, 10 * \widehat{k_i^{t-1}}$$

where hats over variables indicate prior estimates, and superscripts indicate the day on which they are estimated. If more than one of these initial starting values produces estimates, we choose the parameters from the model with the highest adjusted R^2. We estimate the exponential model similarly.

The second, more interesting, problem that we encounter during estimation of the logistic outbreak curves is that STATA's `nl` routine may fail to converge. This failure generally occurs in the transition from relatively slow initial growth to subsequent, more obviously exponential growth as the pandemic proceeds. During this phase of the outbreak, the growth in the number of new cases each day is too large to fit a logistic function, i.e., the drop in the growth of new cases necessary to estimate a carrying capacity has not yet occurred. As a result, and as discussed further below, we estimate both exponential and logistic models for each day of the outbreak. In a future draft we will consider an estimation strategy that nests these functions.

Figure 2 provides an example of simulated "actual" cumulative cases and an estimate of the underlying logistic function for 200 days, using equation 1 to simulate actual data.[8] The predicted values use the cumulative path of reported infections up to day 200 to estimate $\widehat{k_i^{200}}$, $\widehat{c_i^{200}}$, and $\widehat{r_i^{200}}$, and thereby generate predicted cases for each day. As indicated in the figure, inflection point of the logistic curve – a crucial moment in the evolution of the outbreak – occurs at the peak of the new cases curve.

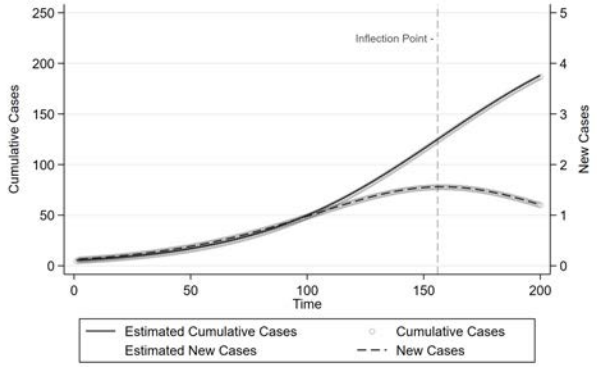
In our application below, we re-estimate the parameters of the exponential and logistic curves each day. That is, for the logistic curve, we estimate $\widehat{k_i^t}$, $\widehat{c_i^t}$, and $\widehat{r_i^t}$ at each day t using the sequence of infections observed up to day $t-1$. To fix ideas, the left panel of Figure 3 illustrates how the logistic parameters evolve over time using the simulated data from Figure 2. As shown in the figure, the estimates in this example are highly volatile in the early stage of the outbreak, are not available due to lack on convergence for days 47 through 78, and then begin to settle down shortly thereafter.

The right panel of Figure 3, by contrast, reports the analogous evolution of the parameters of the exponential estimation. Here, estimates are also volatile in the early days of the pandemic, and settle down near day 50. In contrast to the logistic estimation, parameters are available for each day, i.e., the estimation does not suffer from a lack of convergence. The intuition for the increase in $\widehat{a_{it}}$ and decline in $\widehat{r_{it}}$ as days near 200 is as follows: because the simulated data are logistic, the only way to reconcile them with an exponential function is to assume that the initially exposed $(\widehat{a_{it}})$ is larger, and that the infection spread with a lower growth rate, $\widehat{r_{it}}$.

[7] If the prior day did not converge, we use the most recent prior day for which we have estimates.

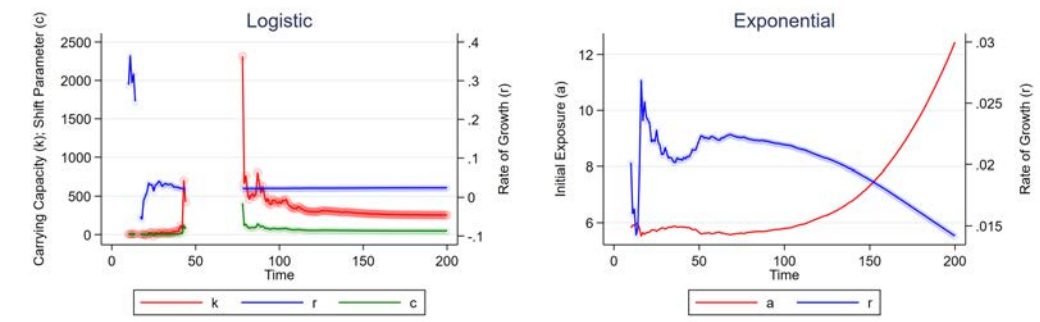
[8] Simulated data are created by computing $C_{it} = \frac{k_i}{1+c_i e^{(-r_i t)}} + |\epsilon_t|$, assuming $k_i = 250$, $r_i = .025$, $c_i = 50$ and $|\epsilon_t|$ is the absolute value of a draw from a standard normal distribution.

Figure 2: Simulated Logistic Pandemic



Source: authors' calculations. Figure compares estimated new and cumulative cases for each day (circles) against "actual" values of those quantities using the simulation procedure noted in the main text. The "actual" data for all 200 days are used to perform the estimation.

Figure 3: Parameter Estimates Using Simulated Logistic Pandemic



Source: authors' calculations. The left panel plots the sequence of logistic parameters, $\widehat{k_{it}}$, $\widehat{c_{it}}$ and $\widehat{r_{it}}$, estimated using the information up to each day t on the simulated data displayed in Figure 2. Right panel of Figure plots the analogous sequence of exponential parameters, $\widehat{a_{it}}$ and $\widehat{r_{it}}$, using the same data. Missing estimates indicate lack of convergence (see text). Circles represent estimates. Solid lines connect estimates.

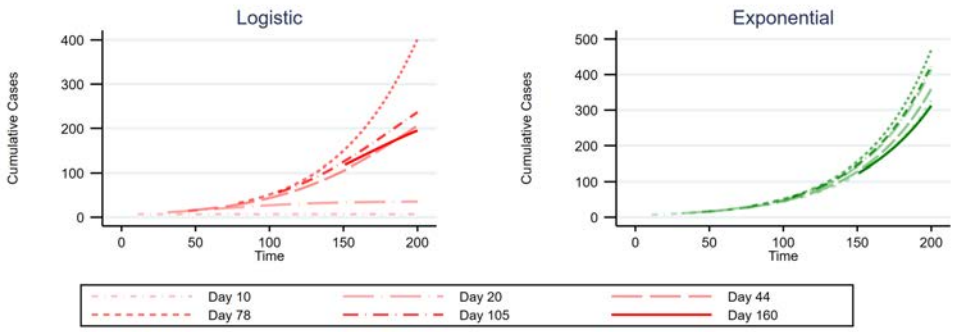
Parameter estimates based on the reported cumulative cases as of day $t-1$ can be used to predict the cumulative number of cases on day t, $\widehat{C_{it}^{t-1}}$, where the superscript $t-1$ refers to the timing of the information used to make the prediction. $\widehat{C_{it}^{t-1}}$ can be compared to the forecast for day t based on the cumulative number of cases as of one day earlier, $\widehat{C_{it}^{t-2}}$. Differences in these predictions capture unexpected changes in severity of the outbreak

Figure 4 compares predicted infections under the logistic (left panel) and exponential (right panel) models using parameters estimated from the cumulative reported infections up to days 10, 20, 44, 78, 105 and 160.[9] In each case, forecasts are given from the day after the last data used until day 200.

As indicated in the figure, early predictions can differ substantially from later predictions. Comparison of the panels in Figure 4 reveals that the parameter estimates, and therefore predicted infections, between days 44 and 78 – i.e., before and after lack of convergence – differ more for the

[9]The logistic prediction for day 44 is the final one available until day 78 due to lack of convergence.

Figure 4: Predicted Cumulative Cases Using Different Days' Estimates (Simulated Data)

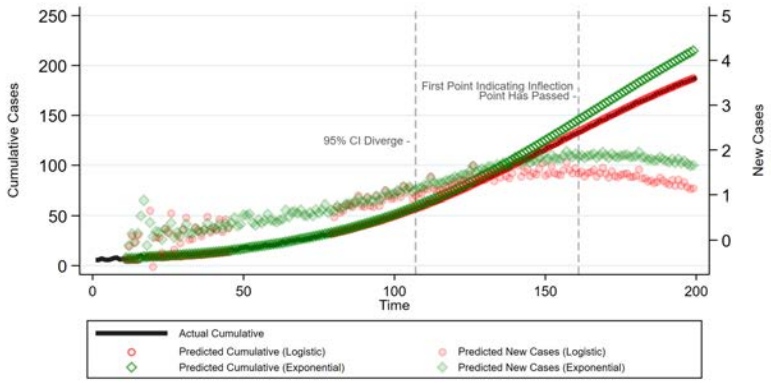


Source: authors' calculations. The right panel plots the predicted sequence of cumulative infections using parameter estimates from the noted day reported in Figure 3. The left panel plots the analogous predictions for the exponential model.

logistic model than the exponential model. More generally, while both sets of estimates exhibit wide variation in the number of cases expected at day 200, this variation is far greater for logistic model estimates.

Finally, Figure 5 compares the logistic and exponential predicted cumulative cases for each day t based on the information available up to day $t-1$, denoted $\widehat{C_{it}^{t-1}}$, where the subscript refers to the prediction day and the superscript refers to the day of the information upon which the prediction is based. The figure also reports analogous predictions for the new cases on each day t, $\widehat{N_{it}^{t-1}}$.

Figure 5: Comparison of Daily Predictions ($\widehat{C_{it}^{t-1}}$) for Simulated Logistic Pandemic



Source: authors' calculations. Figure compares simulated "actual" cumulative infections to predicted infections ($\widehat{C_{it}^{t-1}}$) under the logistic and exponential models. The prediction for each day t is based on the information available up to day $t-1$. The two vertical lines in the figure note when the 95 percent confidence intervals of the two models' predictions initially diverge, and when the logistic model's estimates first indicate that its inflection point has passed.

As illustrated in the figure, $\widehat{C_{it}^{t-1}}$ for the exponential and logistic models line up reasonably well during the initial phase of the pandemic, but begin to diverge at $t = 104$, when the 95 percent confidence intervals for both predictions (not shown) no longer overlap. It is after this point that the logistic model's predictive power begins to exceed that of the exponential model. Indeed, while the exponential model continues to project an ever-increasing number of infections, the logistic

model's predictions head towards the estimated carrying capacity.[10]

2.2 Modeling Economic Impact

Changes in the predictions of the exponential and logistic models of infectious disease described above may be an important input into investors' assessment of the economic impact of a pandemic. For example, a jump in estimated carrying capacity suggests a larger ultimate supply shock in terms of lost labor supply, while an uptick in the estimated growth rate has implications for healthcare capacity constraints.[11]

In our analysis below, we relate information on reported infections to market returns according to the following timing. At the beginning of day t (before markets open) the number of infections occurring on day $t-1$ (released after trading ends on day t) is observed. This day $t-1$ information is used to predict the number of cases for day t, denoted $\widehat{C_{it}^{t-1}}$, where the $t-1$ superscript indicates denotes the day of the information upon which the prediction is based.

In our application below, we compare the change in daily market return, $\Delta ln\left(MV_{it}\right)$, to the log change in the number of predicted cases for day t using information from days $t-1$ and $t-2$,

$$\Delta ln\left(MV_{it}\right) = \alpha + \beta_1 * \Delta ln\left(\widehat{C_{it}^{-2,-1}}\right) + \beta_2 X_{it} + \epsilon_{it} \tag{3}$$

where

$$\Delta ln\left(\widehat{C_{it}^{-2,-1}}\right) = ln\left(\widehat{C_{it}^{t-1}}\right) - ln\left(\widehat{C_{it}^{t-2}}\right). \tag{4}$$

Intuitively, $\Delta ln(\widehat{C_{it}^{-1,-2}})$ captures the unanticipated growth in cases due to a change in the estimated severity of the epidemic. Since both $\widehat{C_{it}^{t-2}}$ and $\widehat{C_{it}^{t-1}}$ are forecasting the cumulative number of cases at time t, the difference between them captures the impact of the new information revealed about the epidemic between $t-2$ and $t-1$. That is, $\Delta ln(\widehat{C_{it}^{-1,-2}})$ is the change in expected cumulative cases due to the updated epidemiological model.[12]

3 Application to SARS

In this section we examine the relationship between changes in infection predictions and aggregate US market returns during the Severe Acute Respiratory Syndrome (SARS) epidemic. According to the World Health Organization, the first SARS case was identified in Foshan, China in November 2002, but was not recognized as such until much later. According to WHO (2006), on February 10, 2003 a member of the WHO in China received an email asking:

> "Am wondering if you would have information on the strange contagious disease (similar to pneumonia with invalidating effect on lung) which has already left more than 100

[10] The separation of the 95 percent confidence intervals of the two models' predictions might be one decision rule that determines a switch from the exponential to the logistic model in real time. Another might be when the logistic model's estimates first indicate that its inflection point has passed. In the logistic model, this point is given by $ln(\widehat{c_{it}})/\widehat{r_{it}}$. It is noted in Figure 5 by the second dashed vertical line.

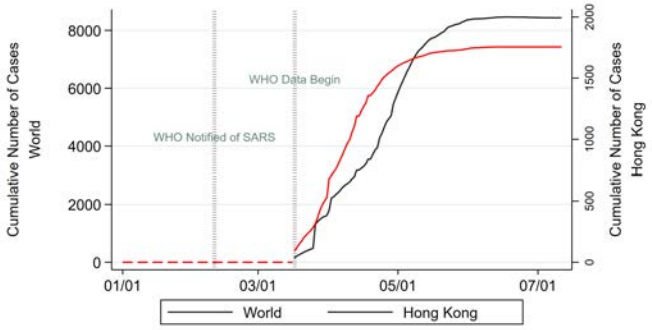
[11] As noted in the introduction, the evolution of these parameters may also trigger policy "events" either directly or as a result of their economic consequences, which may alter the underlying parameters of the outbreak. We do not currently account for such feedback, but plan to do so in a future draft.

[12] We are currently exploring more flexible specifications, e.g., those which might capture the switch between exponential and logistic models, as well as those which reveal any over- or undershooting of reactions.

people dead in ... Guangdong Province, in the space of 1 week. The outbreak is not allowed to be made known to the public via the media, but people are already aware of it (through hospital workers) and there is a 'panic' attitude."

The WHO immediately began and investigation into SARS, and started releasing regular reports of suspected and confirmed cases beginning March 17, 2003.[13] The World Health Organization (WHO) declared SARS contained in July 2003, though cases continued to be reported until May 2004. Figure 6 plots the cumulative number of confirmed SARS infections worldwide (left scale) and in Hong Kong (right scale). The two vertical lines in the figure note the days on which the WHO officially received the aforementioned email, and the first day on which the WHO began reporting the number of infections on each weekday.

Figure 6: SARS Infections in Hong Kong and Worldwide During 2003



Source: World Health Organization and authors' calculations. Figure displays the cumulative reported SARS infections in Hong Kong and the rest of the world from January 1, 2003 to July 11, 2003. The two vertical lines in the figure note the days on which the WHO officially received the aforementioned email, and the first day on which the WHO began reporting the number of infections on each weekday.

Hong Kong and China accounted for the vast majority of cases worldwide.[14] We focus our analysis on Hong Kong for two reasons related to data reliability. First, while China acknowledged having over 300 cases of "atypical pneumonia" in February, the Ministry of Health did not provide day-by-day counts until March 26. In fact, on March 17, the day before WHO began releasing daily situation reports, Chinese authorities informed the WHO that "[t]he outbreak in Guangdong is said to have tapered off." The next day, cases were reported in 8 locations other than China – including Hong Kong. When China did begin reporting daily counts, on March 26, the first count was 800 cases. This large initial level of infections accounts for the sharp jump in World counts displayed for that day in Figure 6. Lack of real-time infection updates prior to this jump undermines reliable estimation of model parameters, thereby impeding accurate assessment of unanticipated changes in infections. Second, it is unclear how China's restrictions on foreign ownership of companies' "A shares" during this period affects the extent to which such unanticipated changes will be reflected in Mainland firms' equity value.

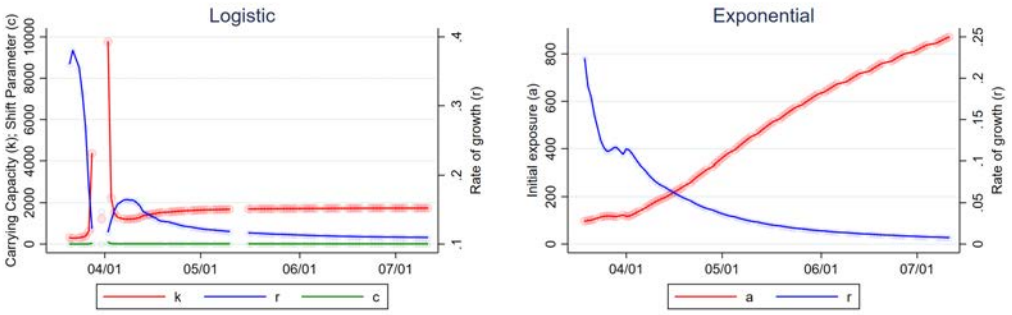
We estimate equations 1 and 2 by day for each country as discussed in Section 2. The daily parameter estimates for the logistic estimation, $\widehat{k_i^t}$, $\widehat{c_i^t}$ and $\widehat{r_i^t}$ are displayed graphically in the left panel of Figure 7. The right panel displays analogous estimates for the exponential function. Gaps

[13] Counts were released every weekday. These data can be downloaded from https://www.who.int/csr/sars/country/en/. A timeline of WHO activities related to SARS events can be found at https://www.who.int/csr/don/2003_07_04/en/.

[14] Reported cases for China are plotted in appendix Figure A.2.

in either panel's time series represent lack of convergence. As indicated in the figure, logistic parameters fail to converge for several days early in the outbreak, and then once again when the estimates have started to settle down in the beginning of May. The exponential model, by contrast, converges on every day in the sample period.

Figure 7: Parameter Estimates for SARS



Source: World Health Organization and authors' calculations. The left panel plots the sequence of logistic parameters, $\widehat{k_{it}}$, $\widehat{c_{it}}$ and $\widehat{r_{it}}$, estimated using the information up to each day t on the cumulative reported cases for Hong Kong displayed in Figure 6. Right panel Figure plots the analogous sequence of exponential parameters, $\widehat{a_{it}}$ and $\widehat{r_{it}}$, using the same data. Missing estimates indicate lack of convergence (see text). Circles represent estimates. Solid lines connect estimates.

In Figure 8, we use these parameter estimates to compare predictions for the two models. Specifically, we use the parameter estimates from day $t-1$ to predict the number of cases under each model for day t, with shading representing the 95 percent confidence interval. As indicated in the panel, predicted infections under the two models (left axis) are similar through the first week in April where they diverge for the remainder of the sample period. Interestingly, this divergence coincides with a stabilization of the estimated inflection point of the logistic curve (right axis), which, as illustrated by the dashed grey line in the panel, hovers between April 5 and 7 from April 5 onward.[15]

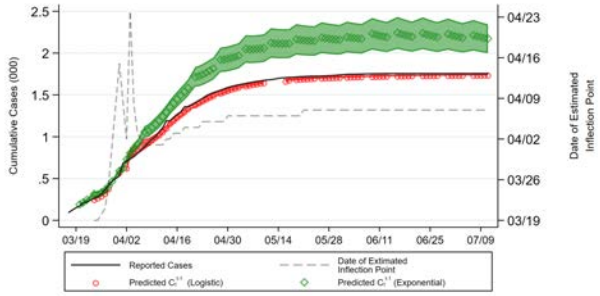
Given our ability to estimate the logistic model for almost all days of the outbreak, we use its estimates in the remainder of our analysis in this section.[16] Figure 9 reports the predicted cumulative cases for day t under the logistic model using information as of day $t-1$, $\widehat{C_{it}^{t-1}}$, and day $t-2$, $\widehat{C_{it}^{t-2}}$, as well as the log difference between these predictions, $\Delta ln(\widehat{C_{it}^{-2,-1}})$. We find that $\Delta ln(\widehat{C_{it}^{-2,-1}})$ exhibits wide swings in value during the early stages of the outbreak, before settling down in late April. As illustrated in Figure A.3, these swings have a noticeably negative correlation with aggregate stock market performance in Hong Kong, as identified via daily log changes in the Hang Seng (downloaded from Yahoo Finance).

We explore this relationship formally in an OLS estimation of equation 3. Coefficient estimates and robust standard errors are reported in Table 1. In the first column, we find a negative and statistically significant relationship using the raw data displayed in Figure A.3. In column 2, we account for weekends and holidays by dividing both the left- and right-hand side variables by the number of days over which the returns are calculated, so that the regression coefficient represents a daily change in market value for a given log change in predicted cases. Here, too, the coefficient estimate is negative and statistically significant at conventional levels, and higher in absolute magnitude.

[15] The inflection point is given by $ln(\widehat{c_{it}})/\widehat{r_{it}}$.

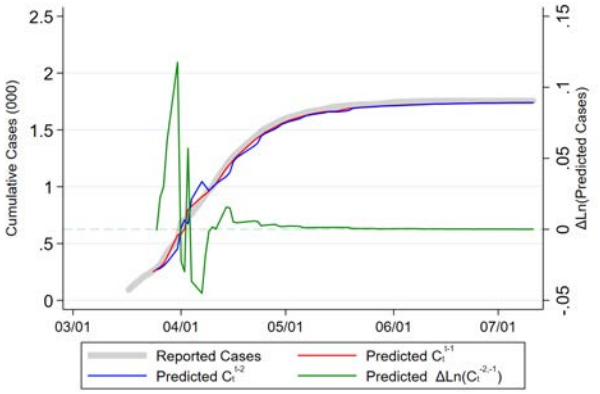
[16] At present, we use the last available estimates for non-convergence days.

Figure 8: Daily Predictions ($\widehat{C_{it}^{t-1}}$) for SARS



Source:World Health Organization and authors' calculations. Figure displays the predicted cases for each day t under the logistic and exponential models using reported cumulative infections as of day $t-1$, using the parameter estimates reported in Figure 7. Solid line tracks reported cases. Shading illustrates predictions' 95 percent confidence intervals. Dashed line (right scale) traces out the estimated day upon which the logistic curve's inflection point ($ln(\widehat{c_{it}})/\widehat{r_{it}}$) is reached. Missing estimates indicate lack of convergence (see text).

Figure 9: Daily Log Difference in SARS Logistic Predictions, $\Delta ln(\widehat{C_{it}^{t-2}})$



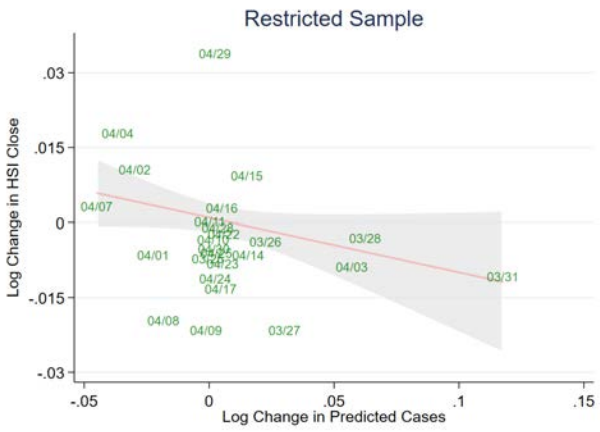
Source: World Health Organization and authors' calculations. Figure reports the predicted cumulative cases under the logistic model displayed in Figure 8 for day t using information as of day $t-1$, $\widehat{C_{it}^{t-1}}$, and day $t-2$, $\widehat{C_{it}^{t-2}}$, as well as the log difference between these predictions, $\Delta ln(\widehat{C_{it}^{-1,-2}})$ and the cumulative reported cases.

In column 3, we examine whether the explanatory power of $\Delta ln(\widehat{C_{it}^{-1,-2}})$ remains after controlling for a simple, local proxy of outbreak severity, the difference in cumulative reported infections between days $t-1$ and $t-2$, $\Delta ln(C_{it}^{-1,-0})$. As indicated in the table, the coefficient of interest remains negative and statistically significant at conventional levels, though of lower magnitude in absolute terms. The coefficient for $\Delta ln(C_{it}^{-1,-0})$ is also negative and statistically significant.

Finally, in column 4, we repeat the specification for column 3 but include month fixed effects to account for potential secular movements in the market unrelated to SARS. Esimate are essentially unchanged.

Overall, the estimates in Table 1 suggests investors may have used simple epidemiological models

Figure 10: Changes in Predicted SARS Cases ($\Delta\widehat{C_{it}^{-2,-1}}$) vs Hang Seng Index Returns



Source: World Health Organization, Yahoo Finance and authors' calculations. Figure displays the daily log change in the Hang Seng Index against the daily log change in predicted cases for day t based on information as of day $t-1$ versus day $t-2$, $\Delta ln(\widehat{C_{it}^{-2,-1}})$.

Table 1: Changes in Predicted SARS Cases vs Hang Seng Index Returns

	(1) ΔLn(Close)	(2) ΔLn(Close)	(3) ΔLn(Close)	(4) ΔLn(Close)
$\Delta Ln(\widehat{C^{-2,-1}})$	-0.0752***	-0.1095***	-0.0891**	-0.0923*
	(0.0241)	(0.0396)	(0.0427)	(0.0537)
$\Delta Ln(C^{-2,-1})$			-0.0445**	-0.0483
			(0.0200)	(0.0294)
Constant	0.0018	0.0010	0.0019*	0.0025
	(0.0013)	(0.0011)	(0.0011)	(0.0051)
Daily Adjustment	N	Y	Y	Y
Month FE	N	N	N	Y
Observations	70	70	70	70
R^2	0.108	0.060	0.103	0.111

Source: World Health Organization,Yahoo Finance and authors' calculations. $\Delta Ln(Close_t)$ is the daily log change (i.e., day $t-1$ to day t) closing values Hang Seng Index. $\Delta ln(\widehat{C_{it}^{-2,-1}})$ is the change in predicted cases for day t using information from days $t-1$ and $t-2$. $\Delta ln(C_{it}^{-1,0})$ is the change in actual observed cases between days $t-1$ and t. Robust standard errors in parenthesis. Columns 2-4 divide all variables by the number of days since the last observation (i.e. over weekends). Column 4 includes month fixed effects.

to update their beliefs about the economic severity of the outbreak in Hong Kong, in real time. Coefficient estimates indicate an average decline of 8 to 11 percent in response to a doubling of predicted cumulative infections.

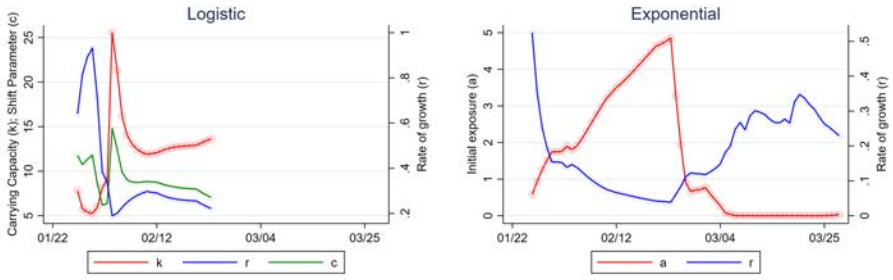
4 Application to COVID-19

In this section we provide real-time estimates of the outbreak parameters and infection predictions for COVID-19 in the United States. We then examine the relationship between changes in these predictions and aggregate equity market returns, as measured by the Wilshire 5000 index downloaded from Yahoo Finance.[17]

Data on the cumulative number of COVID-19 cases in the United States as of each day are from the Johns Hopkins Coronavirus Resource Center.[18] The first COVID-19 case appeared in China in November of 2019, while the first cases in the United States and Italy appeared on January 20, 2020. Our analysis begins on January 22, 2020, the first day that the World Health Organization began issuing situation reports detailing new case emergence internationally. Appendix Figure A.1 displays the cumulative reported infections in the United States from January 22 through March 27, 2020.

We estimate logistic and exponential parameters (equations 1 and 2) for the United States by day as discussed in Section 2.1. The daily parameter estimates for the logistic estimation, $\widehat{k_i^t}$, $\widehat{c_i^t}$ and $\widehat{r_i^t}$ are displayed the left panel of Figure 11, while those for the exponential model, $\widehat{a_i^t}$ and $\widehat{r_i^t}$, are reported in the right panel. Gaps in the time series in either figure represent lack of convergence.

Figure 11: Parameter Estimates for COVID-19



Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. The left panel plots the sequence of logistic parameters, $\widehat{k_{i_t}}$, $\widehat{c_{i_t}}$ and $\widehat{r_{i_t}}$, estimated using the cumulative infections in the US up to each day t. Right panel plots the analogous sequence of exponential parameters, $\widehat{a_{i_t}}$ and $\widehat{r_{i_t}}$, using the same data. Missing estimates indicate lack of convergence (see text). Circles represent estimates. Solid lines connect estimates. Data currently extend to Friday March 27, 2020.

Logistic parameter estimates for the United States fail to converge after February 23, when the number of cases jumps abruptly from 15 to 51. That no parameter estimates are available after this date suggests that growth in new cases observed thus far is inconsistent with a leveling off, or carrying capacity, at least according to our estimation method. The exponential model, by contrast, converges for all days. As a result, we focus on the exponential model for the remainder of our analysis.

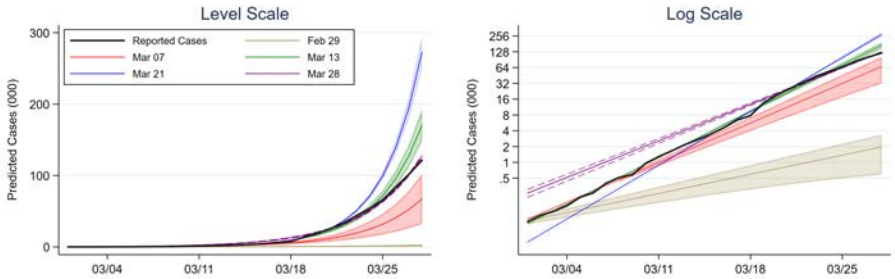
As the sharp changes in US exponential model parameters suggest, predicted cumulative infections vary substantially depending upon the day in which the underlying parameters are estimated. Figure 12 highlights this variability by comparing predicted cumulative infections based on the information available as of February 29 and March 7, 13, 21 and 28. The left panel displays these projections in levels, while the right panel uses a log scale. The five colored lines in the figure trace

[17] We choose this index for its breadth. Results are qualitatively similar for other US market indexes.

[18] These data can be downloaded from https://github.com/CSSEGISandData/COVID-19 and visualized at https://coronavirus.jhu.edu/map.html.

out each set of predictions. Dashed lines highlight 95 percent confidence intervals around these predictions. Finally, the confidence intervals are shaded for all days following the day upon which the prediction is based. To promote readability, we restrict the figure to the period after February 29. The black, solid line in the figure represents actual reported cases.

Figure 12: Predicted Cumulative Cases Using Different Days' Estimates (COVID-19)



Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Figure displays predicted cases for the United States from March 18 onwards using the cumulative reported cases as of five dates: February 29, March 7, March 13, March 21 and March 28. Dashed lines represent 95 percent confidence intervals. Confidence intervals are shaded for all days after the information upon which the predictions are based. Data currently extend to Friday March 27, 2020.

Predicted cumulative infections based on information as of February 29 are strikingly lower than predictions based on information as of March 21 due to the jump in reported cases between those days. Indeed, according to the parameter estimates from March 21, US cases would number close to 300 thousand by the end of March. Equally striking is the downward shift in predicted cumulative cases that occurs between March 21 and March 28. It is precisely these kinds of changes in predicted cumulative cases that our analysis seeks to exploit.

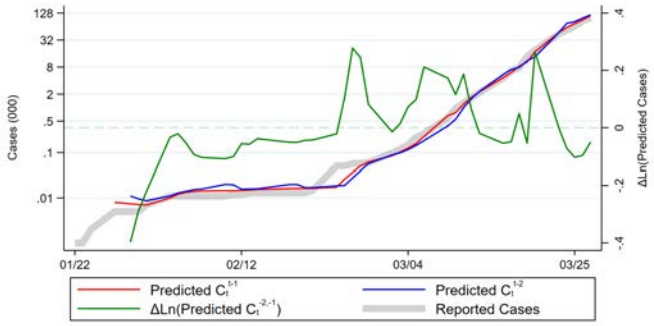
Figure 13 uses the logistic parameter estimates in Figure 11 to plot $\widehat{C_{it}^{t-1}}$ and $\widehat{C_{it}^{t-2}}$ for the exponential model, i.e., the predicted number of cases on day t using the information up to day $t-1$ and day $t-2$. Magnitudes for these cumulative cases are reported on the left axis. The right axis reports $\Delta(\widehat{C_{it}^{-2,-1}})$, the log difference in these two predictions. Intuitively, $\widehat{C_{it}^{t-1}}$ and $\widehat{C_{it}^{t-2}}$ for the most part track each other closely. The former rises above the latter on days when reported cases jump, while the reverse happens when new cases are relatively flat.

Figure 14 plots the daily log change in the Wilshire 5000 index against $\Delta ln(\widehat{C_{it}^{-2,-1}})$. Their negative relationship indicates that unanticipated increases in cases, i.e., $\Delta ln(\widehat{C_{it}^{-2,-1}}) > 0$, are associated with declines in aggregate market value, and *vice versa*. In particular, the approximate 20 percent decline in predicted cases that occurs on March 24 coincides with a greater than 9 percent growth in the market index.

We investigate the relationship displayed in Figure 14 formally by estimating equation 3 via OLS. For each day, we compute $\Delta ln\left(MV_{it}\right)$ as the daily log change in either the closing or opening values of the Wilshire 5000. The estimation period consists of the 47 days from January 22 to March 27. The unit of observation is one day.

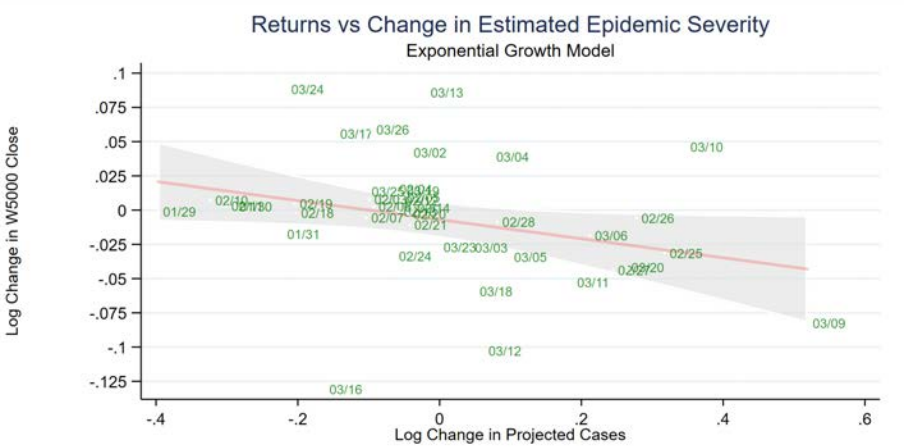
Coefficient estimates as well as robust standard errors are reported in Tables 2 and 3, where the former focuses on the daily opening-to-opening return and the latter on the daily closing-to-closing return. Coefficient estimates in the first column of each table indicate that a doubling of predicted cases using information from day $t-1$ versus day $t-2$ leads to average declines of -7.0 and -3.8 percent for closing and opening prices respectively. These effects are statistically significant at

Figure 13: Daily Logistic Predictions ($\widehat{C_{it}^{t-1}}$ and $\Delta ln(\widehat{C_{it}^{-2,-1}})$) for COVID-19



Source: Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Left axis reports the predicted cumulative cases for day t using information as of day $t-1$, $\widehat{C_{it}^{t-1}}$, and day $t-2$, $\widehat{C_{it}^{t-2}}$, under the exponential model. Right axis reports the log change in these two predictions, $\Delta ln(\widehat{C_{it}^{-2,-1}})$. Data currently extend to Friday March 27, 2020.

Figure 14: Change in Predicted COVID-19 Cases ($\Delta\widehat{C_{it}^{-2,-1}}$) vs Aggregate Market Returns



Source: Johns Hopkins Coronavirus Resource Center, Yahoo Finance and authors' calculations. Figure displays the daily log change in the Wilshire 5000 index against the log change in predicted cases under the exponential model for day t based on day $t-1$ and day $t-2$ information. Data currently extend to Friday March 27, 2020.

conventional levels.

In the second and subsequent columns of each table, we adjust the dependent and independent variables by the number of days since the last trading day. This adjustment insures that changes which transpire across weekends and holidays, when markets are closed, are not spuriously large compared to those that take place across successive calendar days. As indicated in the second column of each table, relationships remain statistically significant at conventional levels and now have the interpretation of daily growth rates. Here, a doubling of predicted cases per day leads to average declines of 8.6 percent for closing and 4.8 percent for opening prices.

In column 3 of each table, we examine whether the explanatory power of $\Delta\widehat{C_{it}^{-2,-1}}$ remains after controlling for a simple, local proxy of outbreak severity, the most recent change in reported

Table 2: Changes in Predicted COVID-19 Cases ($\Delta\widehat{C_{it}^{-1,-2}}$) vs Market Open Returns

	(1) ΔLn(Open)	(2) ΔLn(Open)	(3) ΔLn(Open)	(4) ΔLn(Open)	(5) ΔLn(Open)	(6) ΔLn(Open)
$\Delta Ln(\widehat{C^{-2,-1}})$	-0.038***	-0.048**	-0.057**	-0.057**	-0.060**	-0.054**
	(0.014)	(0.023)	(0.025)	(0.025)	(0.024)	(0.026)
$\Delta Ln(C^{-2,-1})$			0.014	0.015	0.019	0.009
			(0.030)	(0.029)	(0.029)	(0.032)
$I(\Delta SIndex)$				-0.001		
				(0.023)		
$\Delta Ln(SIndex)$					-0.022	
					(0.051)	
Fiscal Stimulus						0.011
						(0.019)
Constant	-0.008*	-0.005	-0.008**	-0.008**	-0.008*	-0.008*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	41	41	41	41	41	41
R^2	0.082	0.075	0.081	0.081	0.096	0.095
Daily Adjustment	N	Y	Y	Y	Y	Y

Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. $\Delta Ln(Open_t)$ and $\Delta Ln(Close_t)$ are the daily log changes in the opening (i.e., day $t-1$ to day t open) and closing values of the Wilshire 5000. $\Delta ln(\widehat{C_{it}^{-2,-1}})$ is the change in predicted cases. $\Delta ln(C_{it}^{-2,-1})$ is the change in actual observed cases between days $t-2$ and $t-1$. $\Delta ln(C_{it}^{-1,0})$ is the change in actual observed cases between days $t-1$ and t. Robust standard errors in parenthesis. Columns 2-6 divide all variables by the number of days since the last observation (i.e. over weekends). Data currently extend to Friday March 27, 2020.

cases. We use a slightly different variable in each table to account for the timing of the opening and closing returns. For the opening price regressions, we use $\Delta Ln(C^{-2,-1})$ under the assumption that the only information available to predict the opening price on day t is the difference in reported cases from days $t-2$ and $t-1$. For the closing price regressions, however, we use $\Delta Ln(C^{-1,0})$ to informally allow for the possibility that, although day t cases are not officially available until after closing, some information might "leak out" during day t trading.

As indicated in the table, these measures are positive but not statistically significant at conventional levels. Moreover, they have little impact on our coefficients of interest. These results suggest that the primary role local increases in reported cases play in determining market value is through their contribution to the overall sequence of reported infections, manifest in the estimated model parameters.

In the final three columns of Tables 2 and 3 we examine the robustness of our results to including coarse controls for policy. As the COVID-19 pandemic has unfolded in the United States, state and local governments as well as the federal government have undertaken various measures to control its spread and limit the economic burden the disease itself imposes. Enactment of such policies is by definition correlated with the severity of the outbreak, and some of them may be designed to stabilize equity markets, confounding our results.

We consider two controls for policy. The first is a country-level index developed at Oxford University, the Government Response Stringency Index (SIndex), which tracks travel restrictions, trade patterns, school openings, social distancing and other such measures, by country and day.[19]

[19]This index can be downloaded from https://www.bsg.ox.ac.uk/research/research-projects/oxford-covid-19-government-response-tracker.

Table 3: Change in Predicted COVID-19 Cases ($\Delta\widehat{C_{it}^{-2,-1}}$) vs Market Close Returns

	(1) ΔLn(Close)	(2) ΔLn(Close)	(3) ΔLn(Close)	(4) ΔLn(Close)	(5) ΔLn(Close)	(6) ΔLn(Close)
$\Delta Ln(\widehat{C^{-2,-1}})$	-0.070**	-0.086**	-0.096***	-0.098***	-0.098***	-0.096***
	(0.032)	(0.032)	(0.034)	(0.035)	(0.035)	(0.033)
$\Delta Ln(C^{-1,-0})$			0.033	0.042	0.039	0.030
			(0.031)	(0.034)	(0.034)	(0.033)
$I(\Delta SIndex)$				-0.016		
				(0.021)		
$\Delta Ln(SIndex)$					-0.025	
					(0.056)	
Fiscal Stimulus						0.021
						(0.015)
Constant	-0.007	-0.003	-0.008*	-0.008*	-0.008*	-0.009**
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	41	41	41	41	41	41
R^2	0.097	0.105	0.124	0.139	0.132	0.147
Daily Adjustment	N	Y	Y	Y	Y	Y

Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. $\Delta Ln(Open_t)$ and $\Delta Ln(Close_t)$ are the daily log changes in the opening (i.e., day $t-1$ to day t open) and closing values of the Wilshire 5000. $\Delta ln(\widehat{C_{it}^{-2,-1}})$ is the change in predicted cases for day t using information from days $t-1$ anfd $t-2$. $\Delta ln(C_{it}^{-2,-1})$ is the change in actual observed cases between days $t-2$ and $t-1$. $\Delta ln(C_{it}^{-1,0})$ is the change in actual observed cases between days $t-1$ and t. Robust standard errors in parenthesis. Columns 2-6 divide all variables by the number of days since the last observation (i.e. over weekends). Data currently extend to Friday March 27, 2020.

We make use of this index in two ways in columns 4 and 5 of Tables 2 and 3. First, we include an indicator function $I\{\Delta SIndex\}$ which takes a value equal to one if the index changes on day t. Second, we use log changes in the index itself, $\Delta Ln(SIndex)$. As indicated in the tables, neither covariate is statistically significant at conventional levels, and their inclusion has little impact on the coefficient of interest.

Our second control for policy is a coarse measure of fiscal stimulus. This dummy variable is set to one for four days (chosen by the authors) upon which major fiscal policies were enacted. The "Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020", which appropriated 8.3 billion dollars for preparations for the COVID-19 outbreak, was signed into law on March 6. Then, from March 25 to March 27, Congress voted for and the President signed into law the 2 trillion dollar "Coronavirus Aid, Relief, and Economic Security Act." As reported in the table, this dummy variable, too, is statistically insignificant at conventional levels, and exerts no influence on the coefficient of interest.

Policy variables' lack of statistical significance is somewhat puzzling. One explanation for this outcome is that these measures are a function of the information contained in the cumulative reported cases, and therefore retain no independent explanatory power. On the other hand, the various government policies included in the SIndex may have offsetting effects. For example, while social distancing measures might be interpreted by the market as a force that reduces the economic severity of the crisis, they may also be taken as a signal that the crisis is worse than publicly available data suggest. At present, we do not have the degrees of freedom to explore the impact of individual elements of the this index, but plan to do so in a future draft when inclusion of additional countries in the analysis allows for panel estimation.

5 Conclusion

This paper shows that day-to-day changes in the predictions of standard models of infectious disease forecast changes in aggregate stock returns in Hong Kong during the SARS outbreak and the United States during the COVID-19 pandemic. In future updates to this paper, we plan to extend the analysis to other countries and pandemics, and to investigate the link between individual firms' returns and their exposure to public health crises via domestic and international input and output linkages as well as the demographics and occupations of their labor forces.

References

Atkeson, A. (2020, March). What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. Working Paper 26867, National Bureau of Economic Research.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of accounting research*, 159–178.

Barro, R. J., J. F. Ursua, and J. Weng (2020, March). The coronavirus and the great influenza pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity. Working Paper 26866, National Bureau of Economic Research.

Berger, D., K. Herkenhoff, and S. Mongey (2020). A seir infectious disease model with testing and conditional quarantine. *Working Paper*.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies 1*(3), 195–228.

Fama, E. F., L. Fisher, M. C. Jensen, and R. Roll (1969). The Adjustment of Stock Prices to New Information. *International Economic Review 10*.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of financial economics 22*(1), 3–25.

Gormsen, N. J. and R. S. Koijen (2020). Coronavirus: Impact on stock prices and growth expectations. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-22).

Greenland, A., M. Ion, J. Lopresti, and P. K. Schott (2019). Using equity market reactions to infer exposure to trade liberalization. Technical report, Working Paper.

Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character 115*(772), 700–721.

Kermack, W. O. and A. G. McKendrick (1937). Contributions to the mathematical theory of epidemics iv. analysis of experimental epidemics of the virus disease mouse ectromelia. *Journal of Hygiene 37*(2), 172–187.

Keynes, J. (1937). The general theory of unemployment.

Knight, F. (1921). Risk, uncertainty, and profit.

Li, R., S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov2). *Science*.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, Volume 1, pp. 19–46.

Piguillem, F. and L. Shi (2020). The optimal covid-19 quarantine and testing policies. *Working Paper*.

Ramelli, S. and A. F. Wagner (2020). Feverish stock price reactions to covid-19. *Swiss Finance Institute Research Paper* (20-12).

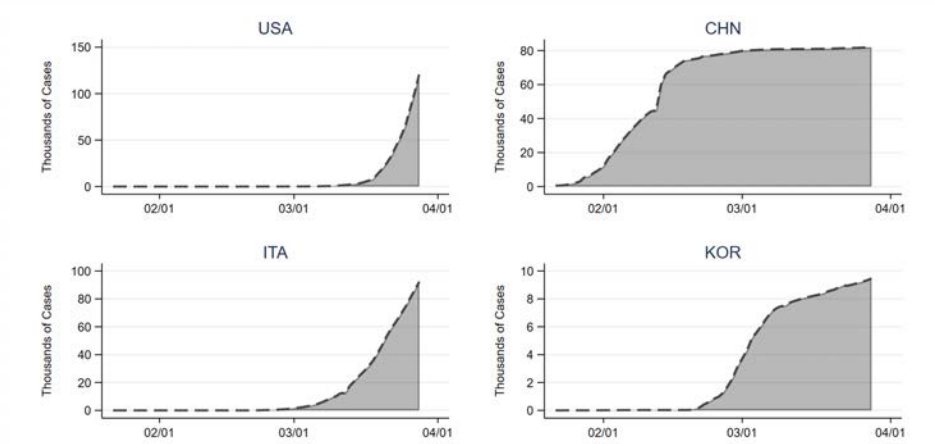
Richards, F. J. (1959, 06). A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany 10*(2), 290–301.

Ross, R. (1911). *The Prevention of Malaria*. John Murray.

Wang, Y.-H., F.-J. Yang, and L.-J. Chen (2013). An investor's perspective on infectious diseases and their influence on market behavior. *Journal of Business Economics and Management 14*(sup1), S112–S127.

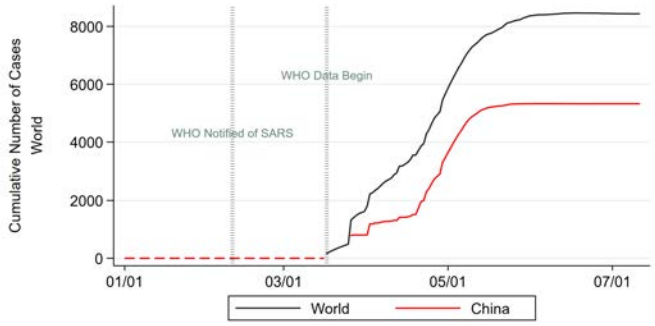
WHO (2006). *SARS: how a global epidemic was stopped*. Manila: WHO Regional Office for the Western Pacific.

Figure A.1: Actual COVID-19 Cases, By Country



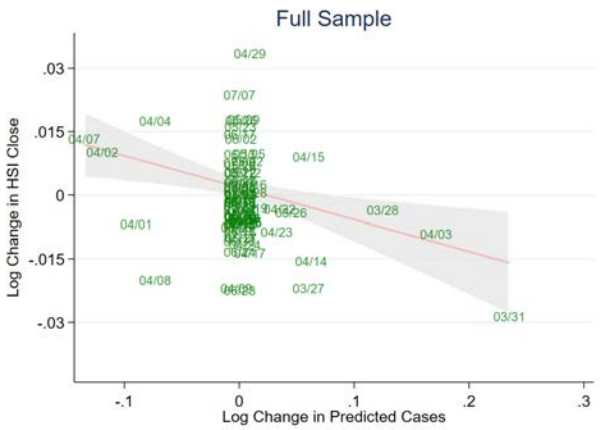
Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Figure displays the COVID-19 up to March 28.

Figure A.2: SARS Infections in China and Worldwide During 2003



Source: World Health Organization and authors' calculations. Figure displays the cumulative reported SARS infections in China and the rest of the world from January 1, 2003 to July 11, 2003.

Figure A.3: Changes in Predicted SARS Cases vs HSI Index



Source: Johns Hopkins Coronavirus Resource Center, Yahoo Finance and authors' calculations. Figure displays the daily log change in the Hang Seng Index against the log change in projected cases for day t based on day $t-1$ and day $t-2$ information.