Aggregate and Firm-Level Stock Returns During Pandemics, in Real Time*

Laura Alfaro† Anusha Chari‡ Andrew Greenland§ Peter K. Schott¶

March 30, 2020

Preliminary and Incomplete – Comments Welcome
COVID-19 analysis based on data through March 27
For most recent version of this paper, click here

Abstract

We model the severity of a pandemic in real time using standard models of infectious disease. We show that changes in these models' predicted infections as a crises unfolds explain day-to-day aggregate market returns. Our analysis currently covers the SARS outbreak and four countries battling COVID-19. Future drafts will extend our investigation to additional countries and pandemics, and examine the relationship between firms' returns and their exposure to pandemics along domestic and international input-output channels, and the demographics and occupations of their workforce.

*This paper is preliminary and incomplete. Missing citations will be be added in a future draft. We thank Nick Barberis, Teresa Fort, Ed Kaplan and Amit Khandelwal for comments and suggestions.
†Harvard Business School & NBER (lalfaro@hbs.edu).
‡UNC Chapel Hill & NBER (achari@unc.edu).
§Martha and Spencer Love School of Business, Elon University (agreenland@elon.edu).
¶Yale School of Management & NBER (peter.schott@yale.edu).

1 Introduction

Pandemics inflict a substantial human toll. They also roil economies and asset markets. In this paper, we show that day-to-day changes in the estimated parameters of standard models of infectious disease as an outbreak unfolds forecast stock returns. This relationship is consistent with investors' using these models to gauge the economic severity of the outbreak, in real time.

We use information on the cumulative number of infections reported each day to predict infections one day ahead assuming the outbreak follows either exponential or logistic growth. We compare these forecasts to the same predictions based on the previous day's information, and show that differences in these forecasts predict daily aggregate equity market returns.

Changes in forecasts retain their explanatory power even after controlling for the most recent increase in reported cases, a finding which implies equity markets may recover even while the number of infections is rising, as investors become increasingly certain of the trajectory of the outbreak. In contrast to the latest daily change in reported infections, or even the latest daily acceleration of reported cases, estimated model parameters explicitly predict the eventual number of people that will be infected (known as the "carrying capacity" under the logistic model), and the speed with which that number will be reached.

Our analysis currently examines aggregate US equity market returns during the SARS and COVID-19 pandemics. In subsequent drafts, we will extend the analysis to other pandemics, and to other countries fighting COVID-19. We also will investigate the link between firm-level returns and firms' exposure to pandemics via domestic and international input and output linkages, as well as various other factors, including the demographics and occupations of their workforces.

This paper contributes to several literatures. First, we add to the very large body of research on asset pricing that examines the predictability of stock returns.[1] In this paper we draw upon epidemiological models to infer how investors might update their beliefs about disease progression.

Second, our (upcoming) 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 events to understand market dynamics. In a typical event study, researchers choose events that are thought to coincide with substantial changes in investor beliefs, and analyze the behavior of firms' abnormal returns relative to an asset pricing model around these events.[2] Here, we (will) show that changes in epidemiological models' predictions regarding the harshness of an outbreak can reveal firms' exposure to this shock via various channels.[3]

Third, our paper contributes to the very large literature in public health which attempts to explain the trajectory of cases during an infectious disease outbreak.[4] In contrast to that research, we link changes in the estimated parameters and predictions of these models in real time to economic outcomes. An interesting question for further research is the extent to which feedback from predicted economic consequences 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.

[1]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.

[2]Bianconi et al. (2018), for example, find that industries and firms subject to greater import competition with China exhibited relatively high stock returns after President Trump's March 22, 2018 memorandum signifying the start of a "trade war" between the US and China. Closer to the context of this paper, Wang et al. (2013) examine 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).

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. Ramelli and Wagner (2020) provide an overview of how the US stock market has evolved since the emergence of COVID-19. Gormsen and Koijen (2020) 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. Sections 3 and 4 apply our framework to SARS and 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,

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

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:

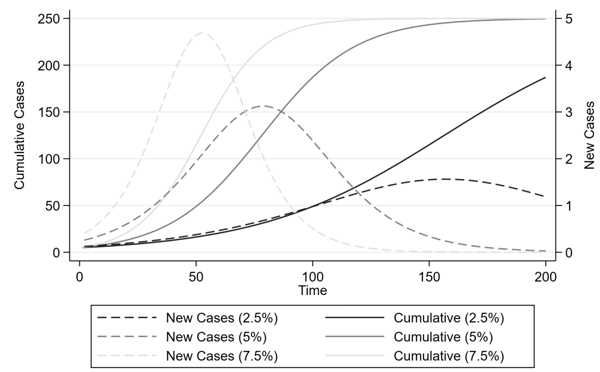
$$C_{it} = \frac{k_i}{1 + c_i e^{(-r_i t)}}, \tag{2}$$

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.

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`).[5] STATA's `nl`

[5] 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).

Figure 1: Disease Outbreak 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).

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, final 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.

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.[6] In the case of the logistic model, we then conduct a grid search defined by all triples $\{r, c, k\}$ such that

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

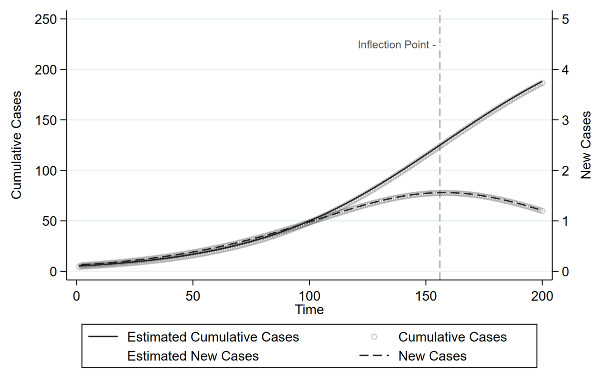
where hats over variables indicate prior estimates, and superscripts indicate the day on which they are estimated. If more than one of these initial starting values produces estimates, we choose the parameters from the model with the highest adjusted R^2. We estimate the exponential model similarly.

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 growth initially to an obviously exponential pattern over time. We believe this problem reflects the fact that, during this phase of the outbreak, the growth in the number of new cases each day is too large to be captured by 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, and assume that investors switch between them once their predictions become sufficiently distinct.

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

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.[7] The predicted values use the cumulative information as of 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. The inflection point of the logistic cumulative cases curve – a crucial moment in the evolution of the outbreak – occurs at the peak of the new cases curve.

Figure 2: Disease Outbreak Simulation



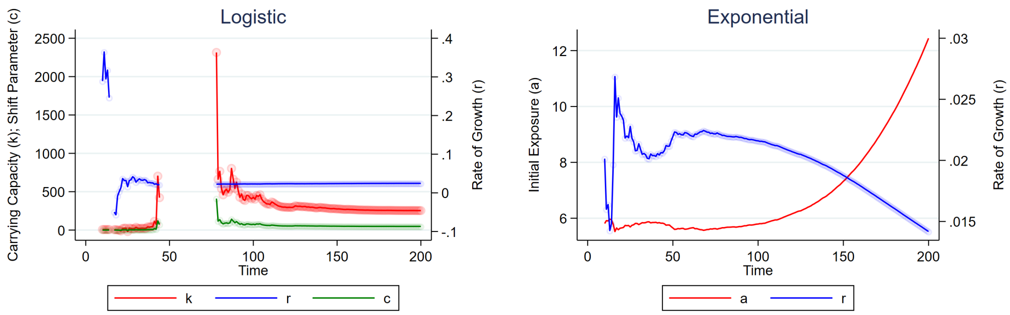
Source: authors' calculations. Figure compares estimated new and cumulative cases for each day (circles) against "actual" values of those quantities using the simulation procedure noted in the main text. The "actual" data for all 200 days are used to perform the estimation.

In our applications 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 $t-1$. 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, in this example, estimates 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 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]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 3: Logistic Parameter Estimates During Simulation

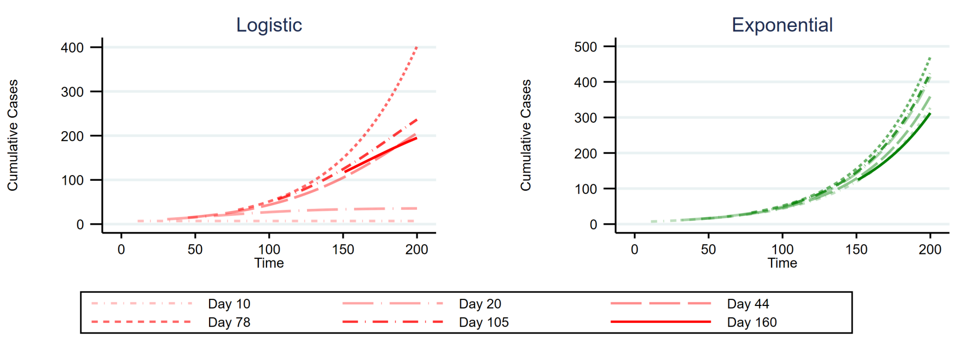


Source: authors' calculations. The left panel plots the sequence of logistic parameters, $\widehat{k_{it}}$, $\widehat{c_{it}}$ and $\widehat{r_{it}}$, estimated using the information up to each day t on the simulated data displayed in Figure 5. Right panel Figure plots the analogous sequence of exponential 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.

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. This prediction can then be compared to one generated for day t based on the cumulative number of cases as of one day earlier, $\widehat{C_{it}^{t-2}}$.

To fix ideas, Figure 4 compares predicted infections from up to day 200 under the logistic model(left panel) and using parameters estimated based on cumulative reported infections from days $t \in \{32, 38, 53, 73, 92, 199\}$. In each case, predictions are displayed for all days after the information upon which they are based.

Figure 4: Logistic Predicted Outbreak Profiles Estimated At Different Dates

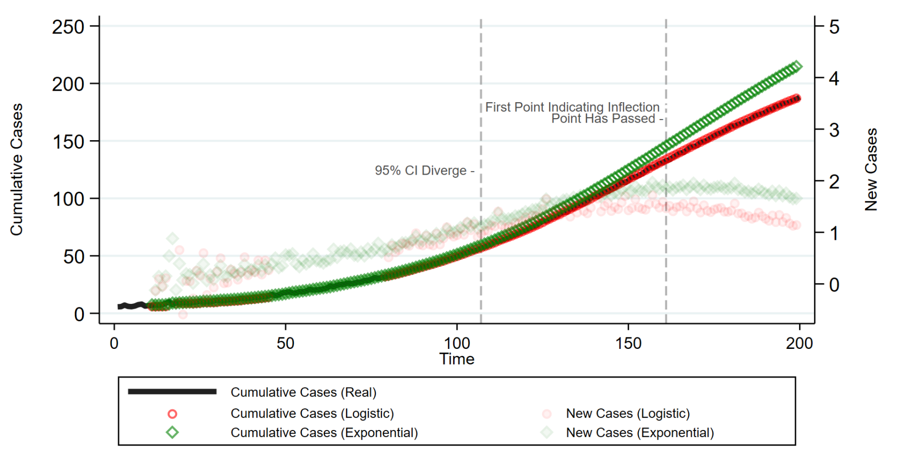


Source: authors' calculations. The right panel plots the predicted sequence of cumulative infections using parameter estimates from the noted day reported in Figure 3. The left panel plots the analogous predictions for the exponential model.

As indicated in the figure, early predictions can differ substantially from later predictions. The prediction for day 44 is the final one available until day 78 due to lack on convergence. Comparison of the panels in Figure 4 reveals that the change in parameter estimates and predicted infections between days 44 and 78, i.e., before and after lack of convergence – are far more distinct for the logistic model (left panel) than the exponential model (right panel). As indicated in these figures, both sets of estimates exhibit wide variation in the number of cases expected at day 200.

Finally, Figure 5 compares the exponential (green) and logistic (red) predictions for cumulative and new cases for each day t based on the information available up to day $t-1$.[8] As illustrated in the figure, the exponential and logistic series line up very well through the initial phase of the pandemic, but begin to diverge at $t = 104$, when the 95% confidence intervals for both predictions 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.

Figure 5: Comparison of Daily Predictions for Logistic and Exponential



Source: authors' calculations. Figure compares simulated "actual" cumulative infections as of day 200 to the predicted infections using parameters estimated on day 200.

While separation of the 95 percent confidence intervals of the two models' predictions might be one decision rule that is used to 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, i.e., when new cases are at their highest – 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.

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.[9]

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 $t-1$ 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.

[8]As discussed further in the next section, we assumed investors compare these predictions in real time in assessing the economic consequences of the pandemic.

[9]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.

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

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

where

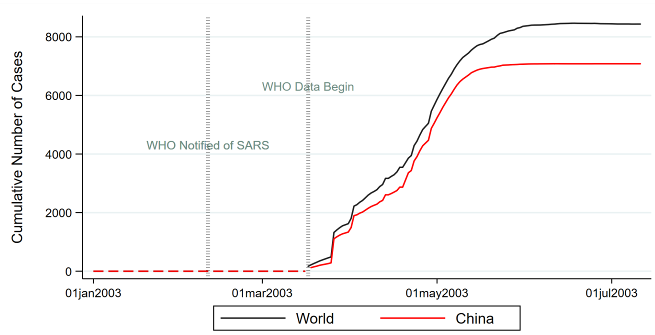
$$\Delta ln\left(\widehat{C_{it}^{-1,-2}}\right) = ln\left(\widehat{C_{it}^{t-1}}\right) - ln\left(\widehat{C_{it}^{t-2}}\right). \tag{4}$$

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. Our expectation is that if $\Delta(\widehat{C_{it}^{-1,-2}}) > 0$, the market will fall.[10]

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. After receiving an email at their Beijing location in February of 2003, the WHO begin investigating SARS and releasing regular reports of suspected and confirmed cases beginning March 17, 2003. The World Health Organization (WHO) declared SARS contained in July 2003, though cases continued to be reported until May 2004.

Figure 6: SARS Infections in China and Worldwide



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.

Data on the cumulative number of SARS infections worldwide are from the WHO website and are summarized in Figure 6.[11] The two vertical lines in the figure note the days on which the WHO

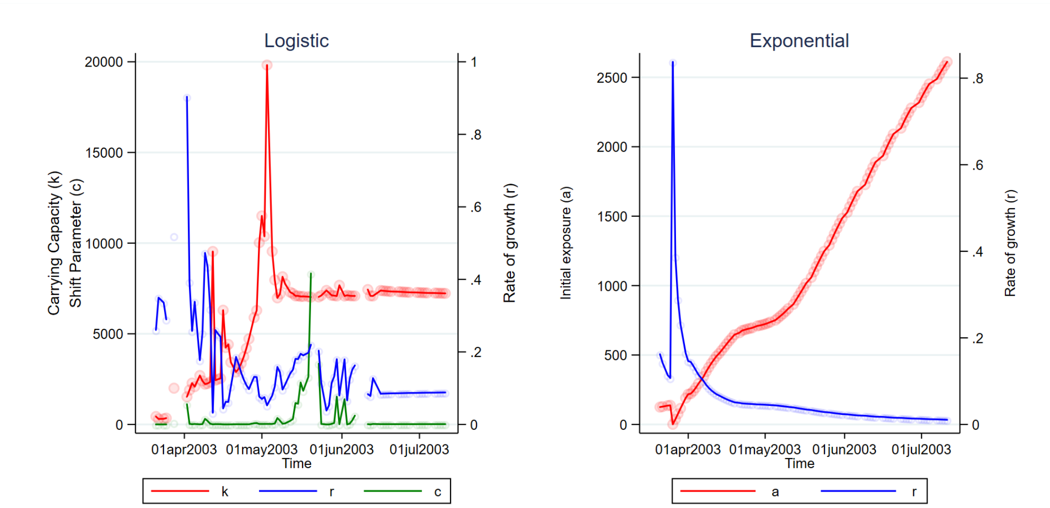
[10] We are currently exploring more flexible specifications, e.g., those which might capture the switch between exponential and logistic models.

[11] Infection data can be downdloaded 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/`.

officially was notified of the SARS outbreak by Chinese authorities, and the first day on which the WHO began reporting the number of infections on each weekday. As indicated by the bold line in the figure, China accounted for the vast majority of cases worldwide. As a result, we focus on estimating the exponential and logistic parameters for China.

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 in the time series in either figure represent lack of convergence.

Figure 7: SARS Parameter Estimates



Source: World Health Organization 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 information up to each day t on the simulated data displayed in Figure 5. Right panel Figure 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.

As indicated in the figure, logistic parameters fail to converge one several days early in the outbreak and then once again once the estimeates have started to settle down in the beginning of May, when there is a brief uptick in cases (see Figure 6). The exponential model, by contrast, converges on every day in the sample period.

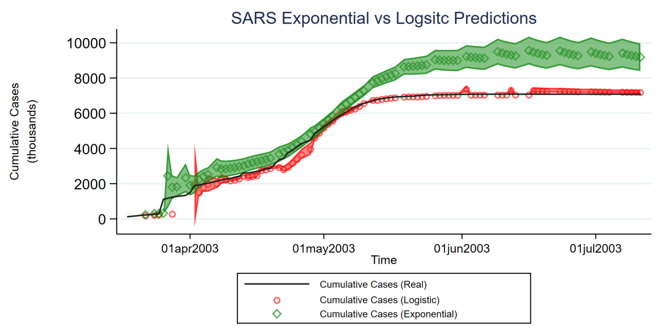
In Figure 8, we use these parameter estimates to compare predictions for the two models. Specifically, we use the parameter estimates from day $t-1$ to predict the number of cases under each model for day t, with shading representing the 95 percent confidence interval. As indicated in the figure, predicted infections under the two models are similar until late April, where they diverge for about a week, and then again in early May, when they begin diverging for good.

We examine the relationship between changes in expected cumulative infections in *China* and aggregate stock market performance in the *United States* in Figure 9.[12] In the figure, the daily log change in the Wilshire 5000 index is plotted against $\Delta ln(\widehat{C_{it}^{-1,-2}})$. The right panel displays the full sample. For better resolution of the overall trend, the right panel of the figure restricts the sample to days in which the log change in predicted cases is between -0.5 and 0.5.

As indicated in the figure, days with larger upward revisions in projected cases have larger

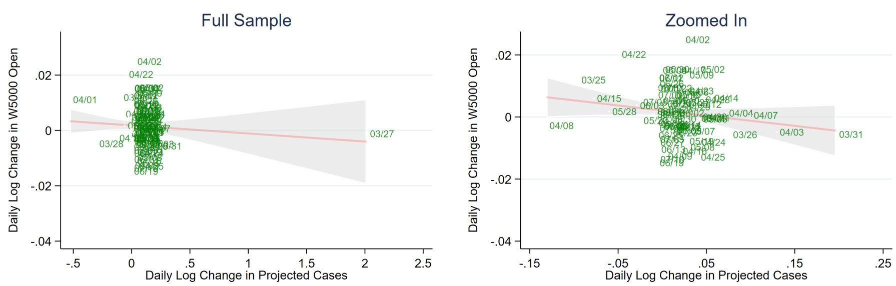
[12] We will include results for equity returns in China and Hong Kong in a future draft. We use the Wilshire 5000 due to its breadth of coverage; results are qualitatively similar for other US market indexes.

Figure 8: SARS Daily Case Predictions



Source: 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. Missing estimates indicate lack of convergence (see text).

Figure 9: Changes in Predicted SARS Cases vs Aggregate US 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 projected cases for day t based on day $t-1$ and day $t-2$ information. Data currently extend to **Friday March 27, 2020**.

downward movements in returns. An interesting feature of this figure is the large vertical cluster of points with relatively small changes in predicted cases. This trend likely is driven by the fact that the SARS epidemic was largely confined to China, and therefore not a dominant influence on the US equity market, which was on an upward trend throughout this pandemic. The secular rise in the Wilshire 5000 during 2003 is displayed in appendix Figure A.2.[13]

[ADD SARS REGRESSION TABLE AND DISCUSSION]

4 Application to COVID-19

In this section we provide real-time estimates of the outbreak parameters and infection predictions for COVID-19 in China, South Korea, Italy and the United States. We then examine the relationship between changes in these predictions and aggregate equity market returns in the United States using equation 3. We will include similar estimations for China, South Korea and Italy in a future draft.

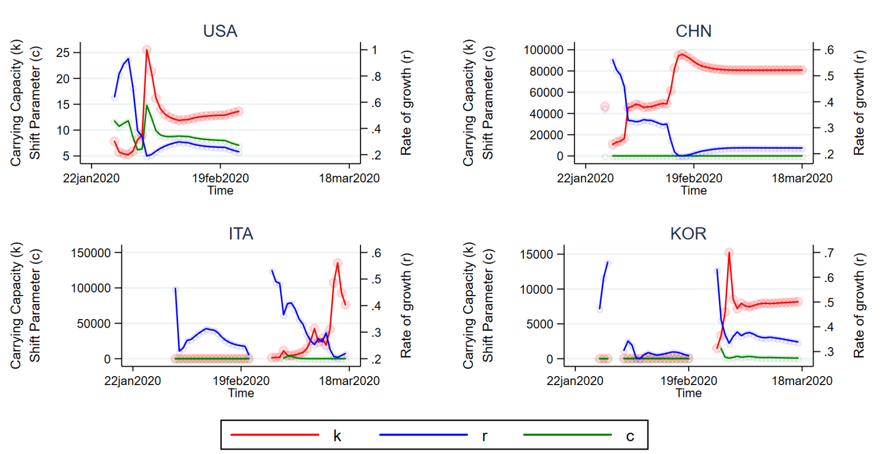
[13] By contrast, COVID-19 is arguably the most substantial current influence on the aggregate US stock market.

Data on the cumulative number of COVID-19 cases in each country as of each day are from the Johns Hopkins Coronavirus Resource Center.[14] As above, for the daily aggregate stock market performance of the United States, we use data on the Wilshire 5000 index from Yahoo Finance.[15]

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. All of our analysis, however, begins on January 22, 2020, the first day that the World Health Organization began issuing situation reports detailing new case emergence internationally. The number of cases in each country across our sample period are displayed in Appendix Figure A.1. Our estimates currently include the cumulative number of cases through Friday March 27, 2020.

We estimate logistic and exponential parameters (equations 1 and 2) by day for each country 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 in Figure 10, while those for the exponential model, $\widehat{a_i^t}$ and $\widehat{r_i^t}$, are reported in Figure 11. Gaps in the time series in either figure represent lack of convergence. For now, we focus our discussion on the parameter estimates for the United States.

Figure 10: Logistic Parameter Estimates



Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Figure displays the estimated growth rates and carrying capacities using observed cumulative cases up to each day. Missing estimates indicate lack of convergence (see text). Circles represent estimates. Solid lines connect estimates. Data currently extend to Friday March 27, 2020.

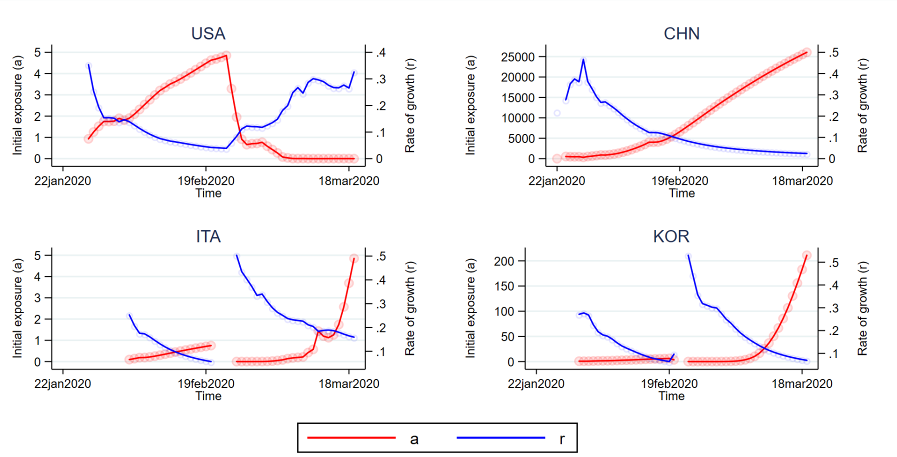
Logistic parameter estimates for the United States fail to converge beginning on February 24, when the number of cases jumps abruptly from 15 to 51. That no parameter estimates are available after this date suggests that growth in new cases observed thus far is inconsistent with a leveling off, or carrying capacity, at least according to our estimation method. The exponential model, by contrast, converges for all days.

As the sharp changes in US exponential model parameters suggest, predicted cumulative infections vary substantially depending upon the particular day of the estimates. Figure 12 highlights this volatility by comparing predicted cumulative infections based on the information available on February 29, March 7, March 13, March 21 and March 28. The five shaded areas in the figure trace out the 95 percent confidence intervals for predicted cumulative infections based on the estimates for each of these days. To promote readability, we restrict the figure to the period after March 18.

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

[15] We choose the Wilshire 5000 for its breadth; results are qualitatively similar for other US market indexes.

Figure 11: Exponential Parameter Estimates

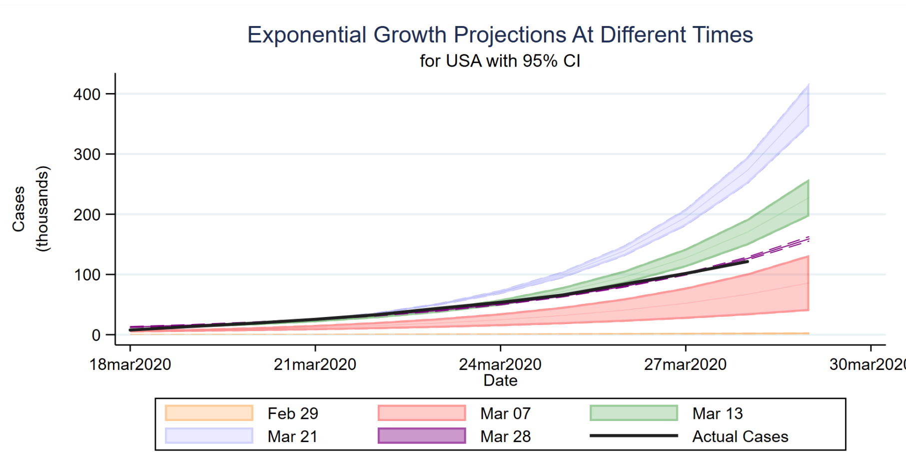


Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Figure displays the estimated growth rates and carrying capacities using observed cumulative cases for each day t using reported cases up to that day. Missing estimates indicate lack of convergence (see text). Circles represent estimates. Solid lines connect estimates. Data currently extend to Friday March 27, 2020.

The black, solid line in the figure represents the actual observed cases.

[MAKE THE MARCH 28 SOLID PURPLE, NOT DASHED, LIKE THE OTHERS]

Figure 12: Exponentially Forecasted Cases Based on Data Available at Different Points.

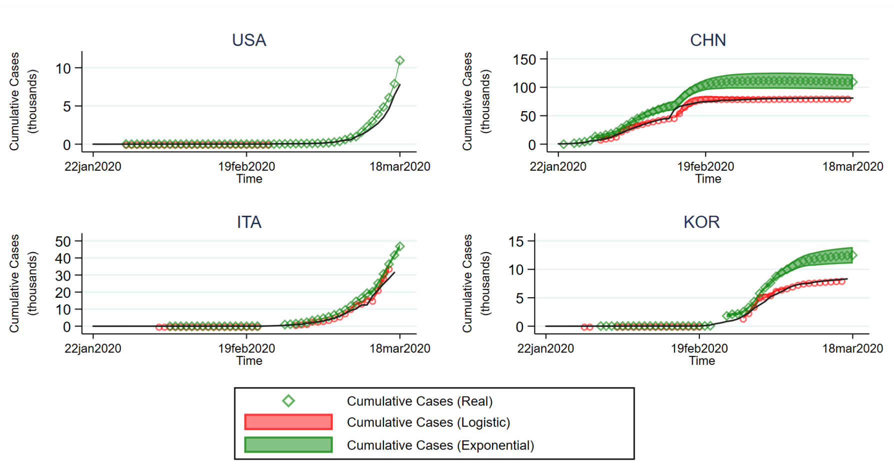


Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Figure displays predicted cases for the United States from March 18 onwards using the cumulative reported cases as of five dates: February 29, March 7, March 13, March 21 and March 28. Shading represents 95 percent confidence intervals. Data currently extend to **Friday March 27, 2020**.

Prediction cumulative infections based on information as of February 29 (orange band) are strikingly lower than predictions based on information as of March 21 (blue band), 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 400,000 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 parameter estimates in Figures 10 and 11 to plot $\widehat{C_{it}^{t-1}}$ for each model, i.e., the predicted number of cases on day t using the information up to day $t-1$. The covariate of interest in our regression analysis below, $\Delta(\widehat{C_{it}^{-1,-2}})$, is the difference between successive points along each curve. As logistic predictions are not available for the United States after February 24 due to lack of convergence, we use the predictions of the exponential model in our examination of equity returns.

Figure 13: Predicted Cases for Exponential and Logistic



Source: Source: Johns Hopkins Coronavirus Resource Center and authors' calculations. Figure displays the projected cases according to both the logistic and exponential models and provides a 95% confidence interval. Predictions are based on the parameter estimates from the prior day. Data currently extend to Friday March 27, 2020.

Figure 14 plots the daily log change in the Wilshire 5000 index against $\Delta ln(\widehat{C_{it}^{-1,-2}})$.[16] The negative relationship displayed in the figure indicates that when new information suggests predicted cases are larger than the prediction of the day before, i.e., when $\Delta ln(\widehat{C_{it}^{-1,-2}}) > 0$, the aggregate market index declines, and *vice versa*. In particular, the greater than 9 percent growth in the market index on March 24 coincides with a reduction in predicted cases of approximately 20 percent.

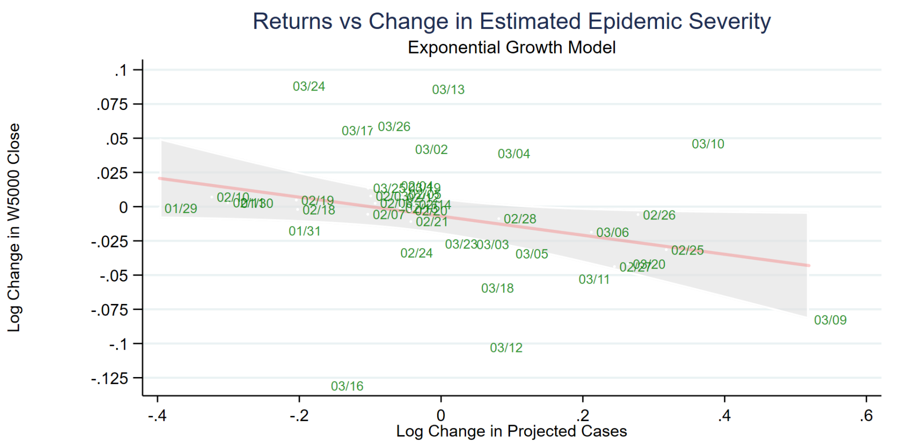
We investigate the relationship displayed in Figure 14 formally by estimating equation 3 via OLS. For each day, we compute $\Delta ln\left(MV_{it}\right)$ as the daily log change in either the closing or opening values of the Wilshire 5000. The estimation period consists of the 47 days from January 22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

Coefficient estimates as well as robust standard errors are reported in Tables 1 and 2, 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 leads to an average declines of -7.0 and -3.8 percent for closing and opening prices respectively. These effects are statistically significant at conventional levels.

In the second column 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 that 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

[16] We will include results for China, South Korea and Italy in a future draft.

Figure 14: Changes in Predicted Cases 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 projected cases for day t based on day $t-1$ and day $t-2$ information. Data currently extend to **Friday March 27, 2020**.

Table 1: Changing Exponential Predictions vs Market Open Returns: USA

	(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. $\Delta ln(\widehat{C_{it}^{-1,-2}})$ 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.

an average decline 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 the change in predicted cases remains after controlling for a simple, contemporaneous proxy of outbreak severity, the most recent, or local, change in reported cases. We use a slightly different variable in each table. For

Table 2: Changing Exponential Predictions vs Market Close Returns: USA

	(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** (0.032)					
$\Delta Ln(\widehat{C^{-2,-1}})$		-0.086** (0.032)	-0.096*** (0.034)	-0.098*** (0.035)	-0.098*** (0.035)	-0.096*** (0.034)
$\Delta Ln(C^{-1,-0})$			0.033 (0.031)	0.042 (0.034)	0.039 (0.034)	0.033 (0.032)
$I(\Delta SIndex)$				-0.016 (0.021)		
$\Delta Ln(SIndex)$					-0.025 (0.056)	
Fiscal Stimulus						0.002 (0.020)
Constant	-0.007 (0.007)	-0.003 (0.005)	-0.008* (0.004)	-0.008* (0.004)	-0.008* (0.004)	-0.008* (0.004)
Observations	41	41	41	41	41	41
R^2	0.097	0.105	0.124	0.139	0.132	0.124

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. $\Delta ln(\widehat{C_{it}^{-1,-2}})$ 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.

the opening price regressions, we use $\Delta Ln(C^{-2,-1})$ to reflect the fact 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 account 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 suggests that the primary role local increases in reported cases play in determining market value is through their impact on underlying market parameters. That is, it appears as if markets are most sensitive to local increases in cases when they exert greater influence on the estimated overall severity and timing of the epidemic

As the COVID-19 pandemic has unfolded in the United States, state and local governments have undertaken various measures to control its spread. Imposition of such policies is by definition correlated with the severity of the outbreak, and it is possible that some policies are put into place precisely to stabilize equity markets. As a result, we investigate whether the relationship between changes in model predictions and aggregate returns are robust to the inclusion of two controls for policy.

First, we consider 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.[17] We make use of this index in two

[17] This index can be downloaded from https://www.bsg.ox.ac.uk/research/research-projects/

ways in columns 4 and 5 of Tables 1 and 2. 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 no impact on the coefficient of the covariate of interest.[18]

Second, we include a coarse control for fiscal policy, the dummy variable "Fiscal Stimulus", which is set to one on March 6 and March 25 through 27, corresponding to 4 days on which major legislation was enacted. The "Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020" was signed in to law on March 6. This act appropriated 8.3 billion dollars in funds for preparations for the COVID-19 outbreak. March 25, 26 and 27 are the three days during which Congress voted for and the President signed into law the 2 trillion dollar "Coronavirus Aid, Relief, and Economic Security Act."

As repoted in the table, this dummy variable, too, is statistically insignificant at conventional levels, and exerts no influence on the coefficient of interest.

`oxford-covid-19-government-response-tracker`.

[18]In our current single-country time-series regression we do not have the degrees of freedom to estimate different effects for different policies captured by the index. We note that accounting for government responses to the crisis requires some caution, as it is unclear what the impact of a particular policy might be. For example, while social distancing measures might be beneficial in terms of controlling the spread of the epidemic, they might be interpreted by the market as a force that reduces the economic severity of the crisis, or a signal that the crisis is worse than publicly available data suggest.

5 Conclusion

This paper shows that day-to-day changes in the predictions of standard models of infectious disease can predict changes in aggregate stock returns across four countries currently battling 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.

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.

Bianconi, M., F. Esposito, and M. Sammon (2018). Trade policy uncertainty and stock market performance.

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.

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*.

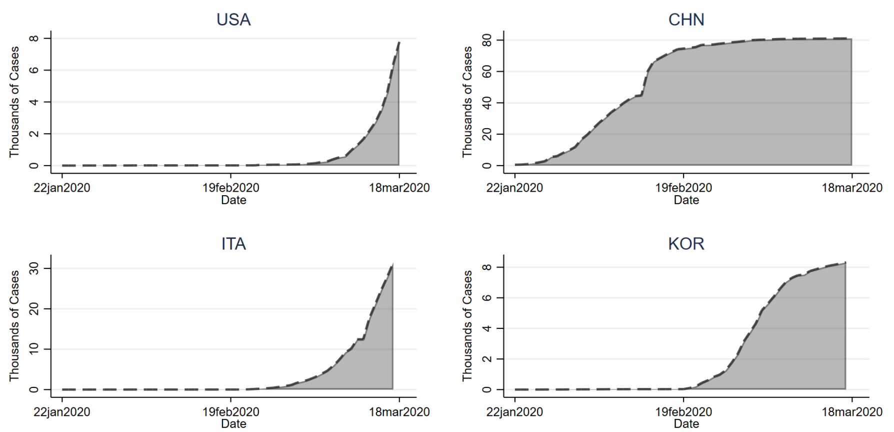
Ramelli, S. and A. F. Wagner (2020). Feverish stock price reactions to covid-19. *Swiss Finance Institute Research Paper* (20-12).

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.

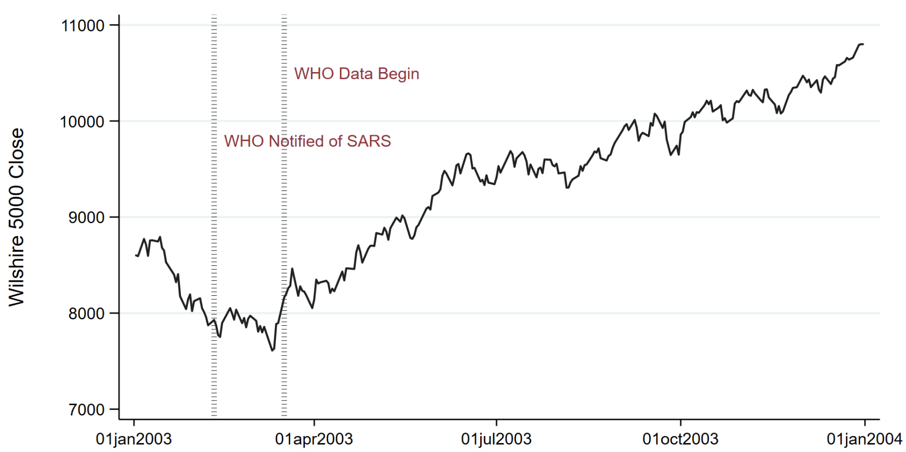
Wang, Y.-H., F.-J. Yang, and L.-J. Chen (2013). An investor's perspective on infectious diseases and their influence on market behavior. *Journal of Business Economics and Management 14*(sup1), S112–S127.

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 22.

Figure A.2: US Wilshire 5000 Index, 2003



Source: Yahoo Finance. Figure displays the Wilshire 5000 aggregate US equity market index.