WINPEC Working Paper Series No.E2107
First draft: April 2021
This version: June 2021

Household Expenditures and the Effective Reproduction Number in Japan: Regression Analysis

Hajime Tomura
Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Household Expenditures and the Effective Reproduction Number in Japan: Regression Analysis

Hajime Tomura*

First draft: April 30, 2021

Updated: June 27, 2021

Abstract

The rate of change in the reported number of new cases of new-coronavirus infection, which approximates the effective reproduction number, is regressed on real household expenditures per household for eating out, traveling, and apparel shopping, as well as mobility in public transportation, using publicly available daily nationwide data in Japan from February 15, 2020, to February 1, 2021. The lagged effects of explanatory variables due to incubation periods are incorporated in the regression. The successful out-of-sample prediction of the effective reproduction number by the regression model indicates that there had been stable correlation between the effective reproduction number and the explanatory variables up to April 2021. The factor decomposition of the fourth wave of infection in April 2021 demonstrates that there were substantial contributions from increases in eating out and traveling as well as mobility related to apparel shopping. Moreover, estimated regression coefficients indicate that real household expenditures for cafe and bar have larger effects on the effective reproduction number per value of spending than the other types of household expenditures in the explanatory variables. Thus, a loss of aggregate demand will be minimized if the effective reproduction number is lowered by restricting household consumption at cafes and bars.

JEL codes: E21, I18.

Keywords: new coronavirus; effective reproduction number; consumption; mobility.

*Faculty of Political Science and Economics, Waseda University, 1-6-1 Nishiwaseda, Shinjyuku-ku, Tokyo, 169-8050, Japan. Email: htomura@waseda.jp. I thank Hiroshi Fujiki, Munechika Katayama, Keisuke Otsu, and seminar participants at various seminars for their valuable comments.

1 Introduction

Household consumption activities have been regarded as part of main causes that spread new-coronavirus infection by generating human-to-human contacts. To quantify this causal relationship, I regress the rate of change in the reported number of new cases, which approximates the effective reproduction number, on real household expenditures per household for eating out, traveling, and apparel shopping, as well as a measure of mobility in public transportation, using publicly available daily nationwide data in Japan. These real household expenditures are included in the explanatory variables because they have been regarded as infectious activities, or the subject of controversial government subsidies, or have shown a high sample correlation with the dependent variable. Because the effective reproduction number measures the number of new cases of infection per an infected person, real household expenditures in the explanatory variables are normalized on a per-household basis. In this paper, I use nationwide data because household expenditure data at daily frequency are available only at national level in government statistics. The sample period of data for the estimation of the regression model is set to the period between February 15, 2020, and February 1, 2021. I do not use more recent data to estimate the regression model because of a possible spread of mutant strains in 2021, which may cause a structural break in the regression. Nonetheless, I use the recent data to see if the regression model can be validated by a good fit of out-of-sample prediction.

In the regression model, the degree of infectious activities on each date is assumed to be a linear function of the aforementioned set of explanatory variables on the same date. Then, the rate of change in the reported number of new cases on each date is assumed to be a weighted sum of past infectious activities over incubation periods, in which the sample distribution of each incubation period from 1 day to 14 days is used as a weight. Here, the sample distribution of incubation periods is interpreted as the probability distribution of incubation periods. In this way, the regression model incorporates lagged explanatory variables without a need to create a new coefficient to estimate for each lag. The model also incorporates time-varying regression coefficients through cross terms between explanatory variables and time dummies for the year-end and new-year holiday period and the states of emergency. The model contains two types of residuals: a white noise for measurement error in the rate of change in the reported number of new cases; and a latent AR(1) process for residual infectious activities on each date. To estimate this model, I use the Bayesian method with an uninformative, or improper, prior distribution for each parameter.

Using the estimated regression model, I generate out-of-sample prediction of the dependent variable from February 2 to May 1 in 2021, given the latest available samples of explanatory variables being up to April 30, 2021, as of the writing of this paper. I will show that out-of-sample prediction traces the realized values of the dependent variable closely. This result indicates that there had been stable correlation between the effective reproduction number and the explanatory variables up to April 2021, and also that the estimates of regression coefficients are not strongly biased.

Given the successful out-of-sample prediction, I decompose a surge in the reported number of new cases in April 2021, i.e., the fourth wave of infection since the onset of the pandemic in Japan, into contributions from explanatory variables in the regression model. It will be shown that an increase in mobility in public transportation had the largest contribution; household expenditure for clothing and footwear had the second largest contribution; and household expenditures for eating out for meals, cafe, bar, and lodging had contributions of similar magnitudes. This result confirms that increases in eating out and traveling contributed to the spread of infection substantially. It also reveals that there was an infectious effect of mobility related to apparel shopping. This effect was not through eating out or the use of public transportation during apparel shopping, because these activities are included as explanatory variables in the regression model separately.

In addition, estimated regression coefficients imply that real household expenditure for cafe has the largest infectious effect per value of spending among household expenditures included as explanatory variables in the regression model. Real household expenditure for bar has the second largest effect. This result implies that a loss of aggregate demand will be minimized if the effective reproduction number is lowered by restricting household consumption at cafes, and then at bars.

This paper is related to the literature on the relationship between mobility and new-coronavirus infection, such as Glaeser, Gorback, and Redding (2020) on U.S. data, and Watanabe and Yabu (2020), Kajitani and Hatayama (2021), and Kurita, Sugawara, and Ohkusa (2021) on Japanese data. Given a high correlation between mobility and household expenditures, the regression analysis in this paper can be interpreted as translating the infectious effect of mobility, which has been confirmed in the literature, into the infectious effect of real household expenditures. The latter measure is useful to discuss economic costs of policy interventions, because it is equivalent to the opportunity cost to contain the spread of new-coronavirus infection in terms of a loss of aggregate demand.

This paper is also related to the large literature on the macroeconomic analysis of the new-coronavirus pandemic. Examples in Japan include Hamano, Katayama, and Kubota (2020), who endogenize a self-restraint on household consumption in an SIR-macro model, and Fujii and Nakata (2021), who combine a reduce-form estimate of the effect of anti-infection social interventions on GDP with an SIR model.[1] While their top-down approaches are useful to endogenize GDP with the spread of infection, this paper takes a bottom-up approach, providing reduced-form estimates of the effects of detailed categories of household expenditures on the spread of infection.

The remainder of this paper is organized as follows: Data sources and the selection of explanatory variables are described in section 2. The regression model is presented in section 3. The estimation of the regression model and the out-of-sample prediction of the effective

[1] For more examples of economic research on the new-coronavirus pandemic in Japan, see the list collected by the Japanese Economic Association at `https://covid19.jeaweb.org/scientific.html`.

reproduction number are reported in sections 4 and 5, respectively. Conclusions are in section 6.

2 Data

The effective reproduction number is determined by the product of three physical factors:[2]

- the rate of effective contact between an infected person and an unimmunized person;
- the probability of infection from an infected person to an unimmunized person per contact; and
- the average period of infection from an infected person.

In this paper, I regress the rate of change in the reported number of new cases, which approximates the effective reproduction number, on a selected set of household expenditures and a measure of mobility to quantify the contributions of household activities to the spread of new-coronavirus infection via effective contacts. In this section, I show the time series of the dependent and explanatory variables, and explain the reasons for the selection of explanatory variables in the regression.

2.1 Data sources

Table 1 summarizes data sources. In this paper, I use the daily estimate of the effective reproduction number published by Toyokeizai-Shinpo-Sha, a publisher in Japan. This estimate is the week-over-week gross rate of change in the reported number of new cases of new-coronavirus infection, raised to the power of 5/7, where 5 is the average generation time (i.e., the average number of days for which an infected person causes the next cohort of infected persons) and 7 is the number of days in the reporting interval to compute the rate

[2] This decomposition is based on a non-technical summary of an SIR model by Suzuki and Nishiura (2020). Note that both the rate of effective contract and the probability of infection from an infected person are affected by social interventions.

of change, which is a week. This simplified formula to estimate the effective reproduction number has been widely used in Japan to update the effective reproduction number real time daily.[3] Also, I use household expenditure data for households with two or more members at national level, because data on nominal household expenditures at daily frequency are publicly available only for this category in the Family Income and Expenditure Survey.

2.2 Sample correlation between the effective reproduction number and nominal household expenditures per household

Figure 1 plots the daily estimate of the effective reproduction number and six types of nominal household expenditures per household: eating out for meals; cafe (including snack accompanying coffee and tea); bar (including meals accompanying alcoholic drink); lodging; domestic travel packages (i.e., bundles of lodging and transportation within the country); and clothing and footwear.[4] Household expenditures in the figure are 7-day backward moving averages, given the aforementioned formula for the daily estimate of the effective reproduction number being an exponential function of the gross rate of change in the reported number of new cases over a week. The sample period starts from March 1, 2020, as the daily estimates of the effective reproduction number are published only from this date.

The first three items are the subcategories of eating out. They have been regarded as infectious activities due to droplets generated by conversations while eating and drinking. As a result, they have been the main subject of government interventions to contain the spread of infection. For example, the government shortened the opening hours of bars and restaurants in populated area during two states of emergency from April 7 to May 25 in 2020

[3]For further discussion on the basis of this formula by Professor Hiroshi Nishiura of Kyoto University, a theoretical epidemiologist, in Japanese, see `https://github.com/contactmodel/COVID19-Japan-Reff` (accessed April 13, 2021).

[4]The original Japanese names of household expenditure variables from the Family Income and Expenditure Survey are such as "Shokuji-dai" for eating out for meals; "Kissa-dai" for cafe; "Inshu-dai" for bar; "Shukuhaku-ryo" for lodging; "Kokunai pakku ryoko-hi" for domestic travel packages; "Hifuku oyobi hakimono" for clothing and footwear. Household expenditure for foreign travel packages is not included here, because it was negligible during the sample period.

and from January 7 to March 21 in 2021, and prohibited the sales of alcohol at bars and restaurants entirely in metropolitan area during the third state of emergency from April 25 to June 20 in 2021. I include these items as explanatory variables in the regression, in order to quantify the contribution of eating out to the spread of infection.

Lodging and domestic travel packages are household expenditure items related to domestic tourism. The government subsidized domestic tourism from July 22 to December 27 in 2020, in order to make up for a loss of revenue for the tourism industry. This subsidy program was called a "Go-To-Travel" campaign. There has been a controversy over whether this campaign helped spreading new-coronavirus infection across the country. I include household expenditure items related to domestic tourism as explanatory variables in the regression, in order to quantify the contribution of domestic tourism to the spread of infection.[5]

Regarding clothing and footwear, this item has been showing a high sample correlation with the daily estimate of the effective reproduction number, as shown in Figure 2. In fact, clothing and footwear has the highest maximum cross correlation coefficient with the daily estimate of the effective reproduction number among large categories of nominal household expenditures, which is as high as the maximum cross correlation coefficient of nominal household expenditure for bar (see Table 2). Even though this observation may be due to eating out during apparel shopping, it may be also due to some independent infectious activities specific to apparel shopping. For this reason, I include household expenditure for clothing and footwear as an explanatory variable in the regression.[6]

[5] Household expenditures for transportation are not included here because they do not separate commutation and tourism. In this regard, it may be possible to include household expenditure for air flights as part of the explanatory variables, because it is rare to commute using air flights. In this paper, however, I choose not to do so, in order to limit the number of explanatory variables, given a small sample size.

[6] Given occasional cluster infections at schools and hospitals, readers may be surprised to see low cross correlation coefficients for education and medical care in Table 2. This observation may be because the values of household expenditures are not closely correlated with the degrees of congestion at schools and hospitals. Perhaps for similar reasons, the maximum cross correlation coefficients for sports club fees (which is part of culture and recreation) and long-term care services (which is part of other consumption expenditures) are as low as 0.27 at lag 4 and 0.13 at lag 17, respectively, for the same sample period as in Table 2. Thus, these

Because there are no other large categories of nominal household expenditures showing high maximum cross correlation coefficients with the daily estimate of the effective reproduction number in Table 2, I will focus on the six household expenditure items shown in Figure 1 as explanatory variables in the regression. This is also due to a need to limit the number of explanatory variables, given the limited length of the sample period since the onset of the pandemic. Later, I will clarify a possible bias in the regression due to an omitted variable problem, and see if a bias is small by reporting the fit of out-of-sample prediction by the regression.

2.3 Sample correlation between the effective reproduction number and mobility

In addition, Figure 3 plots the daily estimate of the effective reproduction number and the six categories of the COVID-19 Community Mobility Reports from Google: retail_and_recreation; transit_stations; grocery_and_pharmacy; workplaces; parks; and residential. Among these, retail_and_recreation, transit_stations, grocery_and_pharmacy, and workplaces can cause human-to-human contacts outside families. retail_and_recreation, however, is closely correlated with nominal household expenditure per household on eating out for meals, as shown in Figure 4. To avoid a multi-collinearity problem, I do not include retail_and_recreation as part of explanatory variables in the regression. Among the remaining three categories of mobility data, transit_stations is included as an explanatory variable representing a general measure of mobility in the regression model, so that the effects of household expenditures are measured separately from the effect of mobility in public places in the regression.

household expenditure items are not included as explanatory variables in the regression. Therefore, cluster infections are captured by residuals in the regression, as will be described below.

3 Regression model

3.1 Definition of variables and the model

Given the discussion described in the previous section, I regress the log of the daily estimate of the effective reproduction number on real household expenditures for eating out for meals, cafe, bar, lodging, domestic travel packages, and clothing and footwear, as well as transit_stations in the COVID-19 Community Mobility Reports from Google. Because the effective reproduction number measures the number of new cases per an infected person, real household expenditures in the explanatory variables are normalized on a per-household basis. Also, the dependent variable is proportional to the rate of change in the reported number of new cases over past 7 days, given the formula for the daily estimate of the effective reproduction number described in section 2.1. Thus, the dependent variable is an observed variable.

Even though a low inflation rate in Japan makes the distinction between nominal and real household expenditures insignificant for most items, household expenditure for domestic travel packages is an exception, because proportional subsidies during the "Go-To-Travel" campaign reduced the net nominal prices of domestic tourism substantially. For this reason, I use real values for all household expenditures in the regression. Real household expenditures per household are computed by dividing nominal household expenditures per household by the corresponding categories of CPI for each, so that their unit is set to 100 yen in their 2020 average prices.[7]

[7]Because only monthly CPI is available, the value of CPI for each month is used for all dates within the same month. The CPI for eating out in general ("Ippan gaishoku" in Japanese) is used to convert nominal household expenditures for eating out for meals, cafe, and bar into real terms, because there is no separate CPI exactly corresponding for each. Because there is no corresponding CPI for domestic travel packages and because the CPI for lodging reflects not only the prices of independent lodging, but also the prices of lodging included in domestic travel packages, I use the CPI for lodging as a proxy to convert nominal household expenditure for domestic travel packages. On the other hand, perhaps because the Go-To-Travel campaign subsidized the costs of both lodging and transportation, nominal household expenditure for domestic travel packages increased substantially during the campaign period, while that for lodging did not, in the Family Income and Expenditure Survey. I linearly interpolate the monthly CPI for lodging between July 2020 and January 2021 to remove the effect of the Go-To-Travel campaign, when I use the CPI for lodging to convert

The form of the regression model is as follows:

$$\ln R_t = \sum_{s=0}^{6}(Z_{t-s} + \eta_{t-s}) \tag{1}$$

$$Z_t = \sum_{k=1}^{14} p_k V_{t-k} \tag{2}$$

$$\begin{aligned} V_t = {} & \alpha_0 + \alpha_1 D_{NY,t} + \alpha_2 D_{AH,t} + \sum_{j=0}^{2} \beta_j D_{SE,j,t} \\ & + \sum_{i=1}^{7}\left[\left(\gamma_i + \delta_i D_{AH,t} + \sum_{j=0}^{2} \phi_{j,i} D_{SE,j,t}\right) X_{i,t}\right] + e_t \end{aligned} \tag{3}$$

$$e_t = \rho e_{t-1} + \epsilon_t \tag{4}$$

where

$$\eta_t \sim N(0, \sigma_\eta^2) \tag{5}$$

$$\epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{6}$$

$$\gamma_i + \delta_i > 0, \quad \gamma_i + \delta_i + \phi_{j,i} > 0 \tag{7}$$

$$\delta_i < 0 \tag{8}$$

$$\rho \in (-1, 1) \tag{9}$$

The initial value of e_t in the estimation, denoted by e_0, is drawn from the unconditional probability distribution for e_t, given (4):

$$e_0 \sim N\left(0, \frac{\sigma_\epsilon^2}{1-\rho^2}\right) \tag{10}$$

The definition of variables is summarized in Table 3.

On the right-hand side of (1) is the sum of Z_{t-s} and η_{t-s} over the past 7 days, including the current date (i.e., for $s = 0, 1, ..., 6$), because the dependent variable on the left-hand side is proportional to the sum of the rate of change in the reported number of new cases over the past 7 days, as described in section 2.1.

nominal household expenditure for lodging in real terms. There is a corresponding CPI for clothing and footwear.

On the right-hand side of (2), p_k for $k = 1, 2, ..., 14$ is the sample distribution of incubation periods in Japan reported by Sugishita, Kurita, Sugawara, and Ohkusa (2020). See Figure 5 for the distribution. To compute the cumulative effect of lagged infectious events on new cases, which is denoted by Z_t, p_k is interpreted as the probability of the incubation period being k days. Then, p_k is multiplied to the degree of infectious events k days ago, i.e., V_{t-k}, for $k = 1, 2, ..., 14$, to measure the contribution from infectious events k days ago to the rate of change in the reported number of new cases on each date. This use of the sample distribution of incubation periods makes it possible to incorporate a relatively long lag length (i.e., 14) without creating a new parameter to estimate for each lag. This is beneficial as the available sample period since the onset of the pandemic is limited.

In (3), the degree of infectious events on each date, V_t, is modeled as a linear function of real household expenditures per household and mobility in public transportation, which are denoted by $X_{i,t}$ for $i = 1, 2, ..., 7$. The other causes of infection are captured by the residual of (3), e_t, which is assumed to have serial correlation, as implied by (4). The residual includes cluster infections not captured by the explanatory variables.[8]

In (3), there are also time dummies for the year-end and new-year holiday period, $D_{NY,t}$, and for the periods before the first state of emergency and during the two states of emergency, $D_{SE,j,t}$ for $j = 0, 1, 2$, as well as a dummy for absolute humidity, $D_{AH,t}$. Through the cross terms between these dummies except $D_{NY,t}$, and $X_{i,t}$ for $i = 1, 2, ..., 7$, (3) incorporates the possibility that the infectious effects of household activities are state-dependent. For the estimation of these effects, (7) imposes restrictions based on a prior expectation that in any state, household activities measured by $X_{i,t}$ for $i = 1, 2, ..., 7$ are infectious to some extent.

To compute $D_{AH,t}$ for each date, a dummy for absolute humidity no less than $9g/m^3$ for

[8]The shock to e_t on each date, which is denoted by ϵ_t, is generated by a normal distribution, as implied by (6). Even though it is possible to assume a distribution with fatter tails for ϵ_t, such as a student-t distribution or a Cauchy distribution, I assume a normal distribution as a benchmark. It turns out that the estimated regression model shows a good fit of out-of-sample prediction, as will be described below. It is left for future research whether the prediction of the model is improved if the model incorporates an alternative distribution for ϵ_t.

the capital of each prefecture is weighted by the population of the prefecture in 2019, and then summed across prefectures to compute the population-weighted nationwide average of the dummies. The threshold level of absolute humidity is set to $9g/m^3$, given the fact that Nottmeyer and Sera (2021) report that the risk ratio of new cases of new-coronavirus infection over absolute humidity was non-linear, and peaked around $6 - 8g/m^3$ in their samples in England. $D_{AH,t}$ approximates such an effect of absolute humidity in each prefecture by a step function. See Figure 6 for the values of $D_{AH,t}$.

A caveat is that the risk ratio is just a sample correlation. Even though, to my knowledge, it is not clear whether there is established evidence for the biological effect of absolute humidity on the infectiousness of new coronavirus, (8) still imposes a negativity restriction on δ_i, i.e., the coefficient to the cross term between $D_{AH,t}$ and $X_{i,t}$, for $i = 1, 2, ..., 7$. This coefficient restriction is based on a prior expectation that at least the infectiousness of new coronavirus does not increase with absolute humidity.

3.2 Sample period

The sample period for the dependent variable is from March 6, 2020, to February 1, 2021. The beginning of the sample period is due to the availability of mobility data from Google.[9] The end of the sample period is set to include explanatory variables only up to January 2021 in the estimation of the regression model. This cap on the sample period is due to a concern on a possible spread of mutant strains in 2021, which may cause a structural break in the regression model. More specifically, the first report on the finding of a mutant strain from an airline passenger from abroad in Japan was on December 18, 2020.[10] By February 10, 2021, 108 cases of mutant strains had been found nationwide.[11] Also, the Tokyo Metropolitan

[9]The COVID-19 Community Mobility Reports from Google are available from February 15, 2020. There are 21 days between the first date of the dependent variable and that of the explanatory variables in the regression, because there are 14-day lags on the right-hand side of (2), and summation over 7 days on the right-hand side of (1).

[10]See `https://www.mhlw.go.jp/content/10900000/000764153.pdf` (accessed on April 14, 2021.)

[11]See `https://www3.nhk.or.jp/news/special/coronavirus/newvariant` (accessed on April 19, 2021.)

Government started screening a sample of PCR-test results to detect mutant strains from December 2020, and found two cases of mutant strains from 1719 samples by January 29, 2021.[12] Thus, the spread of mutant strains was likely to be limited before the end of January 2021.

3.3 Possible biases in the regression model

Before moving on, let me clarify possible biases in the regression model. Among holidays, I only include a time dummy for the year-end and new-year holiday period. This is due to the distinctively different pattern of household behavior during this period, such that mobility in public transportation declines substantially, as can be seen in Figure 3, while people tend to have the largest number of home parties with relatives, which are infectious, in the year. To limit the number of explanatory variables, time-specific infectious events during the other holiday periods, if any, are included in residual infectious events, e_t, in the regression model. This set-up may violate the assumption that e_t follows a homogeneous AR(1) process throughout the sample period.

There is no immediate simultaneous equation bias in the regression model, because all the explanatory variables lag the dependent variable due to incubation periods, as implied by (2). However, if households hold rational expectations of the dependent variable and if the contemporaneous shock to the dependent variable (i.e., η_t) has serial correlation, then

[12] See `https://www.metro.tokyo.lg.jp/tosei/hodohappyo/press/2021/01/30/01.html` (accessed on April 19, 2021.)

the assumption that lagged explanatory variables are uncorrelated with η_t can be violated.[13] Even though, to my knowledge, there exists no household survey to confirm household expectations of future effective reproduction numbers in Japan, rational expectation is a standard assumption in economics.

In addition, because household expenditures and mobility are jointly determined by each household, it is likely that residual infectious activities, e_t, include some household activities that are correlated with explanatory variables in the regression model. Given the difficulty to resolve all the concerns on possible biases with a small sample size and limited data availability, I will later compare out-of-sample prediction of the daily estimate of the effective reproduction number by the regression model with realized data to see if the estimates of regression coefficients are strongly biased.[14]

[13]For illustration, consider the following simple example. Suppose that the effective reproduction number, R_t, is determined by lagged household behavior denoted by x_{t-1}:

$$R_t = \alpha + \beta x_{t-1} + \eta_t$$

where α and β are constant and η_t is an independent white noise. Also suppose that x_{t-1} is determined by the expected value of R_t and other contemporaneous determinants denoted by z_{t-1}:

$$x_{t-1} = \gamma + \theta E_{t-1} R_t + \phi z_{t-1} + \nu_{t-1}$$

where γ, θ, and ϕ are constant and ν_{t-1} is an independent white noise. If households have perfect foresight, then $E_{t-1}R_t = R_t$. In this case, x_{t-1} and η_t become correlated, which is a simultaneous equation bias. If households hold rational expectations, then $E_{t-1}R_t = \alpha + \beta x_{t-1}$. In this case, x_{t-1} remains uncorrelated with η_t. Hence, the presence of rational expectations of future reproduction numbers does not immediately implies that x_{t-1} and η_t are correlated. Nonetheless, if η_t is an AR(1) process, then $E_{t-1}R_t = \alpha + \beta x_{t-1} + \rho\eta_{t-1}$, where ρ is an AR(1) coefficient. As such, households' rational expectations can cause a simultaneous equation bias if x_{t-1} does not incorporate all the structural factors that cause serial correlation of η_t.

[14]To clarify, the current effective reproduction number affects the rate of effective contact between an infected person and an unimmunized person in the future, because it determines the rate of increase in the immunized share of population. Thus, the current effective reproduction number can affect both the effective reproduction number in the future, which is the dependent variable, and the current values of explanatory variables as a confounding factor, causing an endogeneity bias in the regression model. However, given the immunized share of population remaining almost unchanged due to a relatively small number of total cases in Japan, a bias through this channel is likely to be negligible during the sample period.

4 Estimation result

I apply the Bayesian method to estimate parameters in the regression model. I set an uninformative, or improper, prior distribution for each parameter, that is, the density of the prior distribution of each set of parameter values is a constant, given the coefficient restrictions specified by (7) and (8). I use R ver. 4.0.3 (R Core Team 2020) and Rstan ver. 2.21.2 (Stan Development Team 2020) for estimation.[15]

Table 4 shows the posterior mean and the 95% credible interval of each parameter value. The fitted value of the log of the daily estimate of the effective reproduction number and the residuals of the regression model are shown in Figure 7. The fitted value deviates from the observed daily estimate of the effective reproduction number substantially in the summer of 2020 and in November 2020. The bottom panels of the figure imply that these anomalies are mostly due to shocks to residual infectious events, rather than measurement error.

Even though the posterior mean of ϵ_t looks like having serial correlation, the distributions of auto-correlation functions of residuals, i.e., η_t and ϵ_t, in the mcmc samples plotted in Figure 8 imply that serial correlation is mostly removed from residuals by the inclusion of an AR(1) process for residual infectious events, (4), in the regression model.[16]

As shown in Table 4, the posterior means of γ_2 and γ_3, i.e., the coefficients to real household expenditures per household for cafe and bar, respectively, are much larger than the coefficients to the other household expenditures, i.e., γ_i for $i = 1, 4, 5, 6, 7$, in the regression model. Also, the coefficient for cafe is larger than that for bar. This is a natural result, because customers tend to pay more at bars than at cafes, even if they have the same conversations. Thus, the average value of spending per amount of infectious droplets is smaller at cafes than at bars.

[15]The codes and data set for the estimation are available at `https://github.com/hajimetomura/R_HHexp`.

[16]In mcmc sampling, the value of ϵ_t is simulated to compute the likelihood of the value of η_t, i.e., the residual of the observation equation, (1). As a result, the auto-correlation function of ϵ_t is smooth around 0, whereas that for η_t is more fluctuating, as shown in Figure 8.

Because the coefficient to each real household expenditure per household measures the effect of each variable on the effective reproduction number per value of spending, the estimation result shown in Table 4 implies that a loss of aggregate demand will be minimized if the government aims to lower the effective reproduction number by restricting household consumption at cafes, and then bars, by households.[17] Even though the spread of mutant strains and the progress of vaccinations across the population are likely to change the quantitative relationship between the effective reproduction number and the explanatory variables in coming months, if their effects are common across household activities, then the order of regression coefficients described above will be preserved.

A caveat to this result is that household expenditure for cafe in the Family Income and Expenditure Survey does not differentiate spending at regular cafes and that at cafes with karaoke. Because singing is likely to produce more droplets than normal conversations, it is possible that the latter type of cafe drives the estimated regression coefficient to household expenditure for cafe. This issue is left for future research.

5 Out-of-sample prediction of the effective reproduction number from the trough in February 2021 to the fourth wave of infection in April 2021

Using the estimated regression model, I generate out-of-sample prediction of the daily estimate of the effective reproduction number in Japan. The prediction period starts from February 2, 2021, because the estimation of the regression model uses data up to February 1, 2021, as described in section 3.2. The prediction period ends at May 1, 2021, because the samples of explanatory variables are available only up to April 30, 2021, as of the writing

[17] This result is roughly consistent with the fact that, up to the second state of emergency since the onset of the pandemic, the government had been focusing on limiting the opening hours of bars and restaurants up to 8 p.m. in populated area, in order to curb infection through bar consumption at late night. Also, the government aimed to prohibit the sales of alcohol at bars and restaurants entirely in metropolitan area during the third state of emergency from April 25 to June 20 in 2021.

of this paper.[18] Note that the prediction period still includes the fourth wave of infection in April 2021 in Japan, during which there was a surge in the number of new cases of infection across the country.

Figure 9 compares the predicted and realized daily estimates of the effective reproduction number from February 2 to May 1 in 2021, when the time-dummy for the second state of emergency ($D_{SE,2,t}$) is set to zero throughout the prediction period. As shown in the figure, the posterior means of predicted values trace the realized values closely. The good fit of out-of-sample prediction indicates that there had been stable correlation between the effective reproduction number and the explanatory variables in the regression model up to April 2021. It also confirms that the estimates of regression coefficients are not strongly biased.

In addition, it can be shown that if I set the time dummy for the second state of emergency to 1 up to the end of the second state of emergency on March 21, 2021, then the predicted daily estimates of the effective reproduction number would be much higher than the realized values. This result indicates that the declaration of the second state of emergency changed the infectiousness of household consumption and mobility only within January 2021.

Given the successful out-of-sample prediction by the estimated regression model, Figure 10 decomposes changes in the predicted daily estimates of the effective reproduction number from the trough in February 2021 to the peak in April 2021 into contributions of explanatory variables in the regression model. The top panel shows the total contribution of each explanatory variable through both the linear coefficient to the explanatory variable and the coefficient to the cross term between the absolute humidity dummy and the explanatory variable, given the time-dummy for the second state of emergency being set to zero as described above. It indicates that an increase in mobility in public transportation had the largest contribution to the surge in the effective reproduction number during April 2021; household expenditure for clothing and footwear had the second largest contribution;

[18] This is because there is around one-month lag in the release of the Family Income and Expenditure Survey for each month.

household expenditures for eating out for meals, cafe, bar, and lodging had contributions of similar magnitudes; and household expenditure for packaged domestic travels had the smallest contribution.

Even though there have been some public dispute over a need to restrain eating out and traveling to contain the spread of infection, this result confirms that increases in the two types of household consumption contributed to the fourth wave of infection in April 2021 substantially. At the same time, there were also substantial contributions from mobility in public transportation and mobility related to apparel shopping, the latter of which is captured by household expenditure for clothing and footwear.

Even though the infectious effect of household expenditure for clothing and footwear may be surprising to readers, it is the implication of out-of-sample prediction, rather than decomposition based on fitted values. Thus, it is not due to biased estimates of regression coefficients. It is not due to the use of public transportation or eating out during apparel shopping either, because these factors are controlled by separate explanatory variables in the regression model. There can be multiple possible reasons for this result, such as congestion at apparel shops, human-to-human contacts in shopping area where apparel shops are located, or some infectious activities that tend to occur when people go out after buying clothing and footwear. Further investigation on the reason for this result is left for future research.

The bottom panels of Figure 10 separately show the contributions of the explanatory variables via the linear coefficient to each explanatory variable and via the coefficient to the cross term between the absolute humidity dummy and each explanatory variable. The comparison of the two panels implies that an increase in absolute humidity between February and April 2021 had only a minor impact on the result described above.[19]

[19] The contribution of mobility in public transportation in the bottom-right panel of Figure 10 is positive. To understand why, note that the regression coefficient to the cross term between the absolute humidity dummy and each explanatory variable is negative, and also that transit_stations in the COVID-19 Community Mobility Reports from Google measures a decline in mobility in public transportation from the benchmark period from January 3 to February 6 in 2020. Therefore, an increase in absolute humidity reduces the magnitude of a reduction in the dependent variable due to a given decline in mobility in public transportation

In addition to the good fit of out-of-sample prediction described above, the estimated regression model can also replicate existing estimates of the basic reproduction number during an early phase of the pandemic in China, when people were not fully adjusted to the pandemic, if the 2019 data of the explanatory variables are inserted into the model for a hypothetical case of no restriction on household consumption or mobility. See appendix for more details on this result.

6 Conclusions

To quantify the contributions of household activities to the spread of new-coronavirus infection via human-to-human contacts, I regress the rate of change in the reported number of new cases of infection, which approximates the effective reproduction number, on a selected set of real household expenditures per household and a measure of mobility in public transportation, using publicly available daily nationwide data in Japan. The out-of-sample prediction of the daily estimates of the effective reproduction number generated by the estimated regression model closely traces the realized data from the trough in February 2021 to the fourth wave of infection in April 2021. This result implies that there had been stable correlation between the effective reproduction number and the explanatory variables in the regression model up to April 2021, and also that the estimates of regression coefficients are not strongly biased.

The factor decomposition of out-of-sample prediction reveals that the fourth wave of infection in April 2021 were not only due to increases in eating out, traveling, and mobility in public transportation, but also due to an increase in mobility related to apparel shopping, which is captured by household expenditure for clothing and footwear in the explanatory variables. This result is not because people tend to use public transportation or eat out

from the benchmark period. As a result, the contribution of the cross term between the absolute humidity dummy and mobility in public transportation to the dependent variable increases if absolute humidity rises, given the level of mobility in public transportation.

during apparel shopping, since these activities are included as separate explanatory variables in the regression model. It is left for future research to identify the circumstances in which infection is caused by mobility related to apparel shopping.

In addition, the estimates of regression coefficients indicate that a loss of aggregate demand will be minimized if the effective reproduction number is lowered by cutting household consumption at cafes, and then at bars. Even though the spread of mutant strains and the progress of vaccinations across the population are likely to change the quantitative relationship between the effective reproduction number and the explanatory variables in coming months, if their effects are common across household activities, then the order of regression coefficients behind this result will be preserved. Furthermore, if there is an estimate of a change in the probability of infection per contact due to mutant strains and vaccinations, then it can be multiplied with the estimated regression coefficients reported in this paper for adjustment. This issue is left for future research.

A caveat to this result is that household expenditure for cafe used in this paper does not differentiate regular cafes and cafes with karaoke. It is possible that the latter type of cafe drives the estimated regression coefficient to household expenditure for cafe. The differentiation of the two types of cafes is also left for future research.

References

[1] Fujii, Daisuke, and Taisuke Nakata, 2021. "Covid-19 and Output in Japan," RIETI Working Paper 21-E-004, Research Institute of Economy, Trade and Industry. https://www.rieti.go.jp/jp/publications/dp/21e004.pdf (accessed April 28, 2021.)

[2] Glaeser, Edward L., Caitlin Gorback, and Stephen J. Redding, 2020. "JUE Insight: How Much does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities." Journal of Urban Economics. https://doi.org/10.1016/j.jue.2020.103292 (accessed April 28, 2021.)

[3] Hamano, Masashige, Munechika Katayama, and So Kubota (2020) "COVID-19 Misperception and Macroeconomy," WINPEC Working Paper Series No. E2016, Waseda University. https://www.waseda.jp/fpse/winpec/assets/uploads/2020/11/E2016_20201102-1version.pdf (accessed April 28, 2021.)

[4] Imai N, et al., 2020. Report 3 : Transmissibility of 2019-nCoV. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-01-25-COVID19-Report-3.pdf (accessed April 28, 2021.)

[5] Kajitani, Yoshio, and Michinori Hatayama, 2021. "Explaining the effective reproduction number of COVID-19 through mobility and enterprise statistics: Evidence from the first wave in Japan," PLoS ONE 16(3): e0247186. https://doi.org/10.1371/journal.pone.0247186 (accessed April 28, 2021.)

[6] Kurita, Junko, Tamie Sugawara, and Yasushi Ohkusa, 2021. "Effects of climate conditions, mobility trends, and countermeasures on the COVID-19 outbreak in Japan," Medrxiv, https://doi.org/10.1101/2020.12.29.20248977 (accessed April 28, 2021.)

[7] Nottmeyer, Luise N., and Francesco Sera, 2021. "Influence of temperature, and of relative and absolute humidity on COVID-19 incidence in England - A multi-city time-series

study," Environmental Research, 196:1109-77. https://doi.org/10.1016/j.envres.2021.110977 (accessed April 28, 2021.)

[8] R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed April 28, 2021.)

[9] Stan Development Team, 2020. Stan Modeling Language Users Guide and Reference Manual, VERSION 2.21.2. https://mc-stan.org (accessed April 28, 2021.)

[10] Sugishita, Yoshiyuki, Junko Kurita, Tamie Sugawara, and Yasushi Ohkusa, 2020. Effects of voluntary event cancellation and school closure as countermeasures against COVID-19 outbreak in Japan. PLoS ONE 15(12): e0239455. https://doi.org/10.1371/journal.pone.0239455 (accessed April 28, 2021.)

[11] Suzuki, Ayako, and Hiroshi Nishiura, 2020. "COVID-19. Topics: IV. Mathematical modeling and control of infectious diseases," The Journal of the Japanese Society of Internal Medicine, 109:2276-2280. https://www.naika.or.jp/jsim_wp/wp-content/uploads/2020/11/nichinaishi-109-11-article_4.pdf (accessed April 28, 2021.)

[12] Watanabe, Tsutomu, and Tomoyasu Yabu, 2020. "Japan's Voluntary Lockdown", Covid Economics 46(1), pp. 1-31. https://cepr.org/content/covid-economics-vetted-and-real-time-papers-0 (accessed May 7, 2021.)

A Counterfactual simulation with 2019 data

To confirm that the estimates of regression coefficients are not strongly biased by an another exercise, I show a counterfactural simulation of the daily estimates of the effective reproduction number with the 2019 data of explanatory variables, which can be interpreted as a hypothetical case of no restriction on household consumption or mobility. I will compare the simulation result with an independent estimate of the range of the basic reproduction number in literature.

Because the COVID-19 Community Mobility Reports from Google does not exist for 2019, I create an index of mobility in public transportation for 2019 by dividing the monthly average of railway passengers in each month by the monthly average in January 2020. The nationwide average of the monthly number of railway passengers is published in the Statistical Survey on Railway Transport by the Ministry of Land, Infrastructure, Transport and Tourism.

The indexation of railway passenger data is consistent with the feature of the COVID-19 Community Mobility Reports such that each type of mobility data in the reports are expressed in the form of the rate of change from the average over the period between January 3 and February 6 in 2020. Because only the monthly averages of railway passengers are publicly available, I simply use the monthly average in each month for the daily value on each date within the same month. This substitution can be justified by a high correlation between transit_stations in the COVID-19 Community Mobility Reports from Google and the monthly average of railway passengers in 2020, as shown in Figure 11.

Using the 2019 data, I simulate the daily estimates of the effective reproduction number for 365 days from March 6, which coincides with the first date for the daily estimates of the effective reproduction number in 2020-2021 data. For the simulation, I use the data of explanatory variables from February 14, 2019, i.e., one day before the sample period of explanatory variables for the estimation, because 2020 is a leap year.

To simulate the daily estimates of the effective reproduction number over a year, I connect the year end of the 2019 data with the new year data on January 1, 2019, so that the 2019 data loop as hypothetical data without any restriction on household consumption or mobility. Figures 12 and 13 compare the 2019 data of explanatory variables with the 2020-21 data that are used in the estimation of the regression model.

To make comparison between the simulated and observed daily estimates of the effective reproduction number, I only change the values of real household expenditures and mobility in public transportation to the 2019 data in the simulation. I keep using the 2020-21 data for absolute humidity (i.e., $D_{AH,t}$) as well as the dummy for the year-end and new-year holiday period (i.e., $D_{NY,t}$). I set zero to all dummies related to the states of emergency (i.e., $D_{SE,j,t} = 0$ for $j = 0, 1, 2$).

Figure 14 plots the posterior mean and the 95% credible interval of $\ln R_t$ in the simulation with hypothetical 2019 data, along with the observed and the fitted value of $\ln R_t$ for 2020-2021 from March 6, 2020. The figure indicates that without any restriction on household consumption or mobility, the effective reproduction number would have risen around the end of the fiscal year (i.e., the end of March); after the Golden Week holiday period in early May; and in November and December for 2020-2021.

Table 5 summarizes the posterior distribution of annual means of $\ln R_t$ in the simulation with hypothetical 2019 data. Because the 2019 data in the simulation are used for a hypothetical case without any policy intervention or self-restraint, the geometric annual mean of simulated effective reproduction numbers is comparable with the basic reproduction number (i.e., the average number of new cases per an infected person in a population where everyone is susceptible to infection). Indeed, the simulation result shown in Table 5 is largely consistent with the range of existing estimates of the basic reproduction number during an early phase of the pandemic in China between December 2019 and January 2020, when people in the country were yet to be fully adjusted to the pandemic. The range was between 1.4 and 3.5 (see Imai, et al., 2020). This result adds to the out-of-sample prediction described in section 5 to confirm that the estimates of regression coefficients are not strongly biased.

Table 1: Data sources

Data	Level	Frequency	Source
Daily estimate of the effective reproduction number	Nationwide	Daily	Toyokeizai-Shinpo-Sha
Nominal household expenditures per household	Nationwide	Daily	Households with two or more members, Family Income and Expenditure Survey, Ministry of Internal Affairs and Communications
Consumer Price Index (CPI)	Nationwide	Monthly	Ministry of Internal Affairs and Communications
Mobility in public transportation	Nationwide	Daily	transit_stations, COVID-19 Community Mobility Reports, Google
Temperature, Relative humidity	Prefectural	Daily	Japan Meteorological Agency
Populations	Prefectural	Annual	Population estimates, Ministry of Internal Affairs and Communications
Sample distribution of incubation periods	Nationwide	–	Sugishita, Kurita, Sugawara, and Ohkusa (2020)

Table 2: Cross correlation coefficients between the effective reproduction number and 7-day moving averages of nominal household expenditures of large categories per household

	Maximum cross correlation coefficient	Corresponding lag of nominal household expenditures
Food	0.21	10
Housing	0.16	16
Fuel, light and water charges	-0.02	9
Furniture and household utensils	0.25	10
Clothing and footwear	0.62	12
Medical care	0.20	22
Transportation and communication	0.27	12
Education	0.37	5
Culture and recreation	0.39	10
Other consumption expenditures	0.46	8
Bar	0.65	9

Notes: The table shows the maximum cross correlation coefficients between the contemporaneous estimate of the effective reproduction number and lagged 7-day backward moving averages of nominal household expenditures per household. The sample period is from March 1, 2020, to April 30, 2021, as the daily estimates of the effective reproduction number are available only from March 1, 2020.

Table 3: Definition of variables

R_t	Daily estimate of the effective reproduction number
$X_{1,t}$	Real household expenditure per household on eating out for meals
$X_{2,t}$	Real household expenditure per household for cafe
$X_{3,t}$	Real household expenditure per household for bar
$X_{4,t}$	Real household expenditure per household for lodging
$X_{5,t}$	Real household expenditure per household for domestic travel packages
$X_{6,t}$	Real household expenditure per household for clothing and footwear
$X_{7,t}$	transit_stations in the COVID-19 Community Mobility Reports for Japan, nationwide
$D_{SE,0,t}$	Time dummy for the period before the first state of emergency (- 2020/4/6)
$D_{SE,1,t}$	Time dummy for the first state of emergency (2020/4/7-2020/5/25)
$D_{SE,2,t}$	Time dummy for the second state of emergency (2021/1/7-2021/3/21)
$D_{NY,t}$	Time dummy for Dec. 29-Jan. 3.
$D_{AH,t}$	Population-weighted average of the dummy for absolute temperature no less than $9g/m^3$ across the capitals of prefectures.
p_k	A sample distribution of incubation periods in Japan.
V_t	Degree of daily infectious events.
Z_t	Cumulative effect of lagged infectious events on new cases of new-coronavirus infection.
e_t	Residual infectious events.
ϵ_t	Shocks to residual infectious events.
η_t	Measurement error.

Notes: The daily estimate of the effective reproduction number is the week-over-week gross rate of change in the reported number of new cases, raised to the power of 5/7. The unit of each type of real household expenditure per household is 100 yen in the 2020 average price. To compute $D_{AH,t}$ for each date, a dummy for absolute temperature no less than $9g/m^3$ is constructed for the capital of each prefecture, weighted by the population estimate for the prefecture in 2019, and then summed across prefectures to compute the population-weighted nationwide average of the dummies.

Table 4: Estimated regression coefficients

	Posterior mean	2.5%	97.5%		Posterior mean	2.5%	97.5%
α_0	-0.083	-0.186	0.014	ϕ_{01}	-0.001	-0.023	0.023
α_1	0.054	0.002	0.154	ϕ_{02}	-0.003	-0.353	0.355
α_2	-0.018	-0.061	-0.001	ϕ_{03}	-0.024	-0.209	0.151
β_0	-0.100	-0.289	0.072	ϕ_{04}	0.010	-0.068	0.118
β_1	0.073	-0.212	0.384	ϕ_{05}	0.008	-0.047	0.081
β_2	0.220	-0.301	0.807	ϕ_{06}	0.032	-0.015	0.088
γ_1	0.012	0.001	0.032	ϕ_{07}	-0.000	-0.002	0.002
γ_2	0.187	0.025	0.515	ϕ_{11}	0.021	-0.016	0.086
γ_3	0.108	0.013	0.280	ϕ_{12}	0.425	-0.237	1.759
γ_4	0.047	0.007	0.112	ϕ_{13}	0.399	-0.099	1.173
γ_5	0.031	0.004	0.084	ϕ_{14}	0.614	-0.002	1.678
γ_6	0.018	0.002	0.041	ϕ_{15}	0.993	0.077	2.123
γ_7	0.002	0.000	0.005	ϕ_{16}	0.013	-0.025	0.076
δ_1	-0.002	-0.009	-0.000	ϕ_{17}	0.004	-0.001	0.010
δ_2	-0.051	-0.190	-0.001	ϕ_{21}	0.033	-0.014	0.130
δ_3	-0.025	-0.092	-0.001	ϕ_{22}	2.080	-0.033	5.749
δ_4	-0.016	-0.057	-0.000	ϕ_{23}	1.038	-0.050	2.976
δ_5	-0.013	-0.046	-0.000	ϕ_{24}	0.195	-0.036	0.670
δ_6	-0.004	-0.013	-0.000	ϕ_{25}	0.649	0.019	1.585
δ_7	-0.001	-0.003	-0.000	ϕ_{26}	0.029	-0.022	0.118
ρ	0.743	0.346	0.959	ϕ_{27}	0.020	0.002	0.043
σ_η	0.027	0.024	0.029				
σ_ϵ	0.050	0.030	0.088				

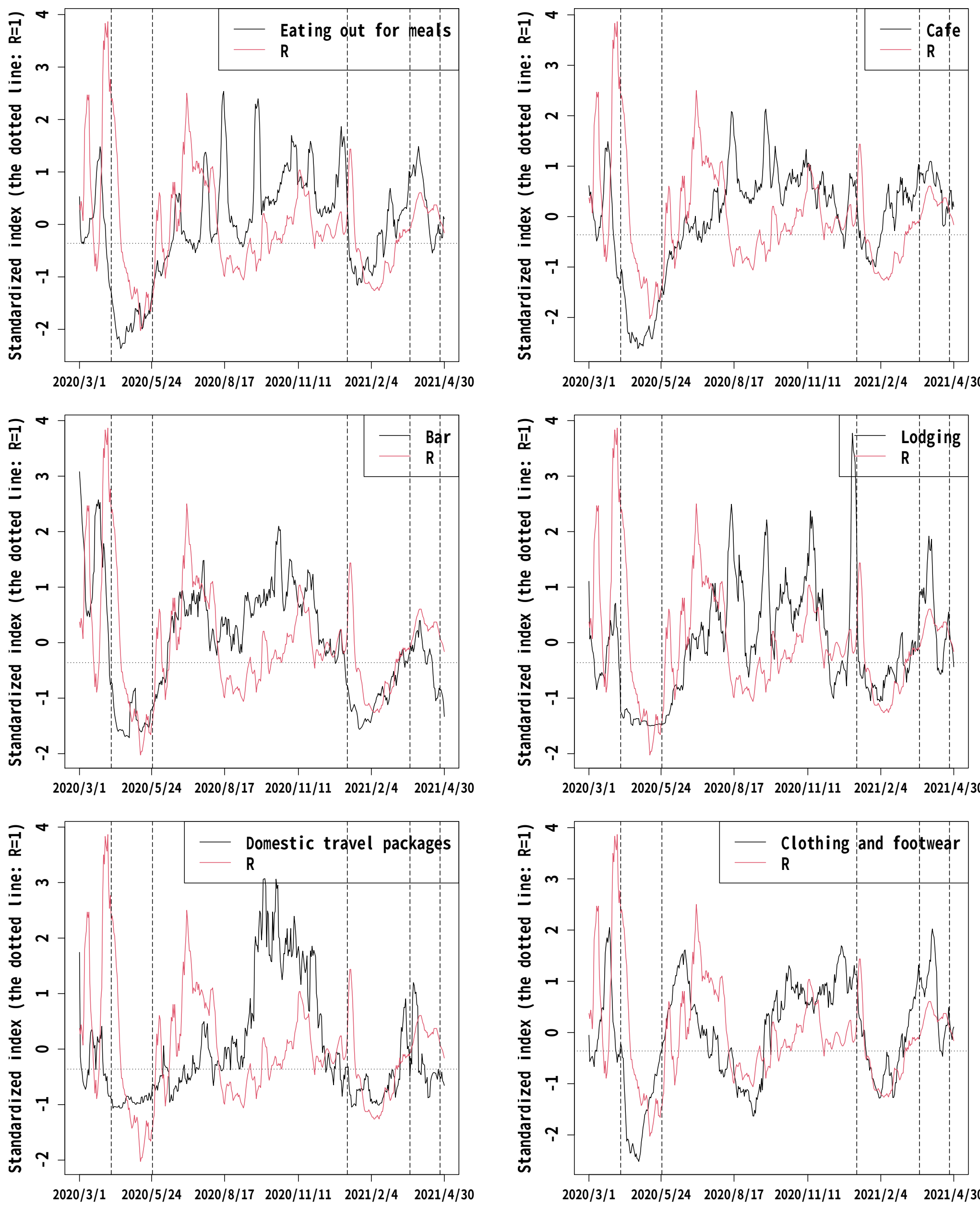
Notes: "2.5%" and "97.5%" indicate the percentiles of mcmc samples. The sample period for the dependent variable is from March 6, 2020, to February 1, 2021. The number of observations is 333. The prior distribution is an improper distribution for each parameter.

Table 5: Posterior distribution of annual means of $\ln R_t$ in the simulation with hypothetical 2019 data

	Posterior mean	2.5% percentile	97.5% percentile
Annual mean of $\ln R_t$	0.94	0.49	1.57
(Corresponding geometric annual mean of R_t)	(2.57)	(1.63)	(4.81)

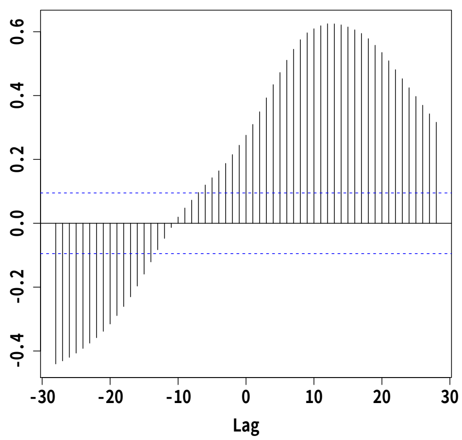
Note: Each cell shows the posterior mean or a percentile of annual means of $\ln R_t$ simulated by inserting the hypothetical 2019 data of real household expenditures and mobility in public transportation into the estimated regression model. In the parenthesis below each figure is the exponential value of the figure, which corresponds to the geometric annual mean of the effective reproduction number implied by the figure.

Figure 1: Effective reproduction number and 7-day moving averages of nominal household expenditures per household



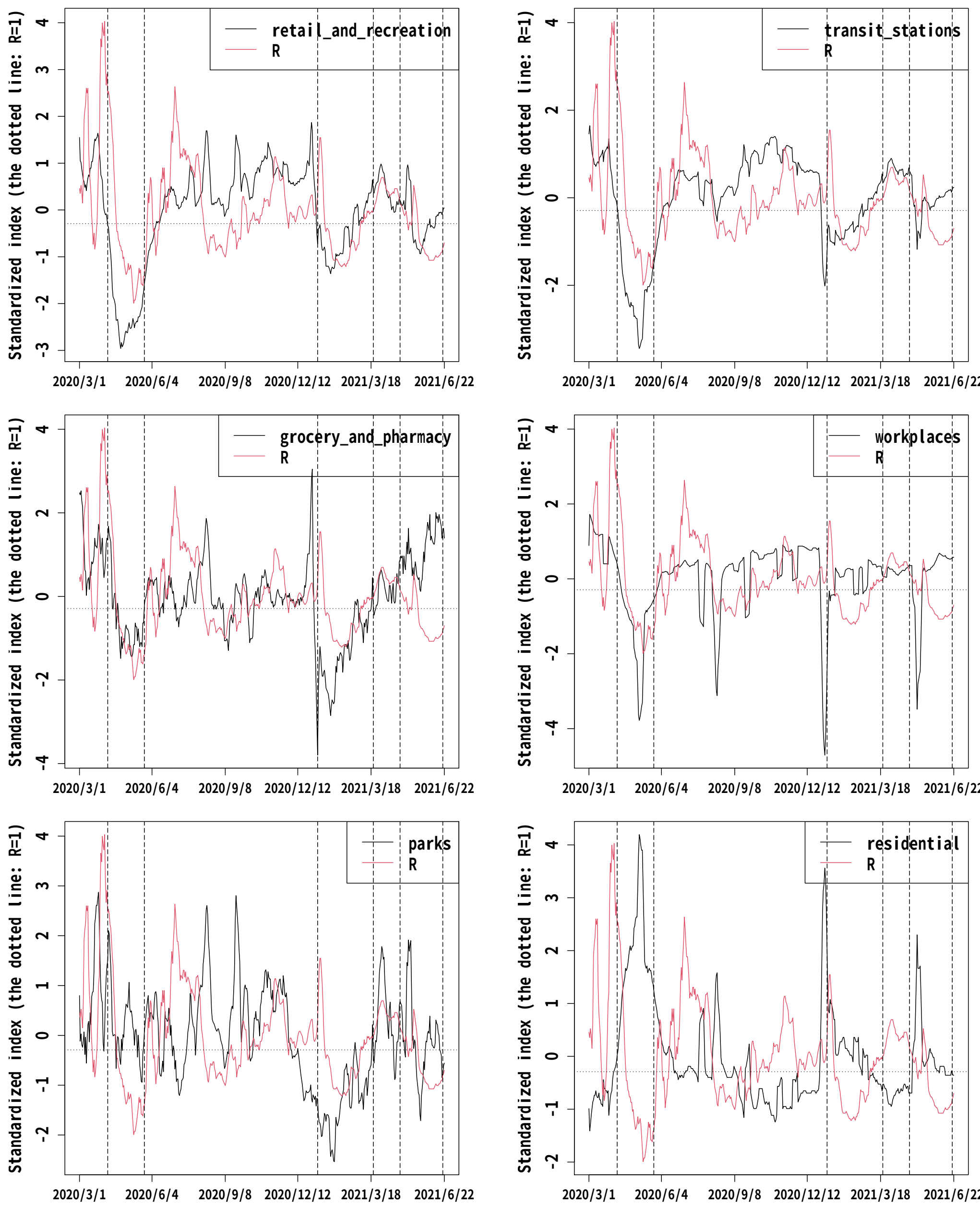
Notes: In each panel, "R" indicates the daily estimate of the effective reproduction number, and nominal household expenditure per household is a 7-day backward moving average. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021, to June 20, 2021. All figures are standardized by their means and standard deviations. The horizontal dotted line indicates the value of the standardized index for the effective reproduction number equal to 1 in each panel.

Figure 2: Cross correlation function between the effective reproduction number and the 7-day moving average of nominal household expenditure for clothing and footwear



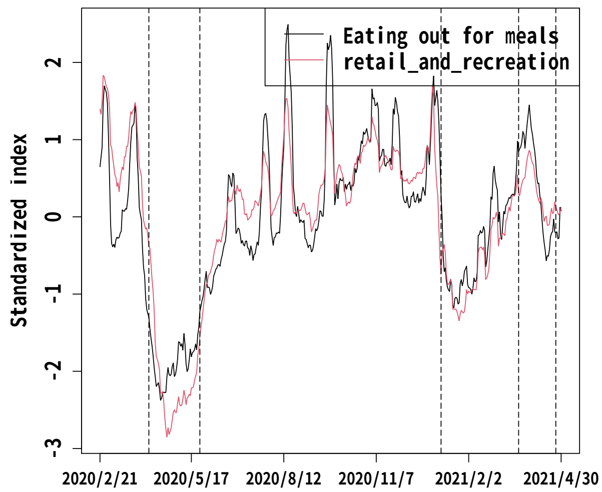
Notes: The figure shows the correlation coefficient between the contemporaneous estimate of the effective reproduction number and lagged 7-day backward moving averages of nominal household expenditure per household for clothing and footwear. On the horizontal axis, negative lags are leads. Horizontal dashed lines are the 95% confidence interval for correlations between independent white noises. The sample period is from March 1, 2020, to April 30, 2021, as the daily estimates of the effective reproduction number are available only from March 1, 2020.

Figure 3: Effective reproduction number and 7-day moving averages of mobility measures



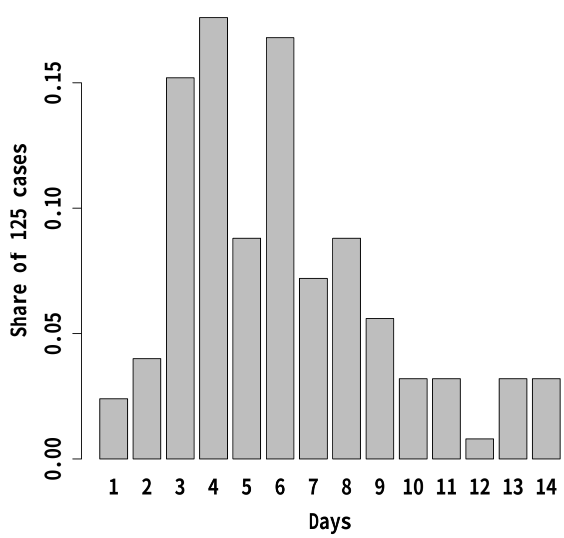
Notes: In each panel, "R" is the daily estimate of the effective reproduction number, and the measure of mobility is a 7-day backward moving average. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021, to June 20, 2021. All figures are standardized by their means and standard deviations. The horizontal dotted line indicates the value of the standardized index for the effective reproduction number equal to 1 in each panel.

Figure 4: 7-day moving averages of mobility in retail and recreation and real household expenditure per household on eating out for meals



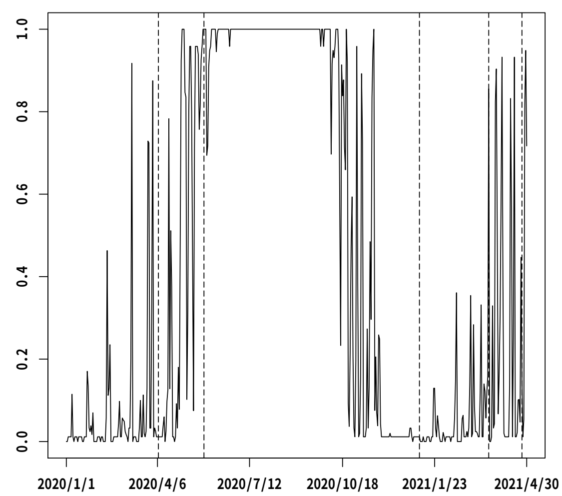
Notes: The figure plots retail_and_recreation in the COVID-19 Community Mobility Reports from Google and real household expenditure per household on eating out for meals. Both figures are 7-day backward moving averages, and standardized by their means and standard deviations. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021, to June 20, 2021.

Figure 5: A sample distribution of incubation periods in Japan



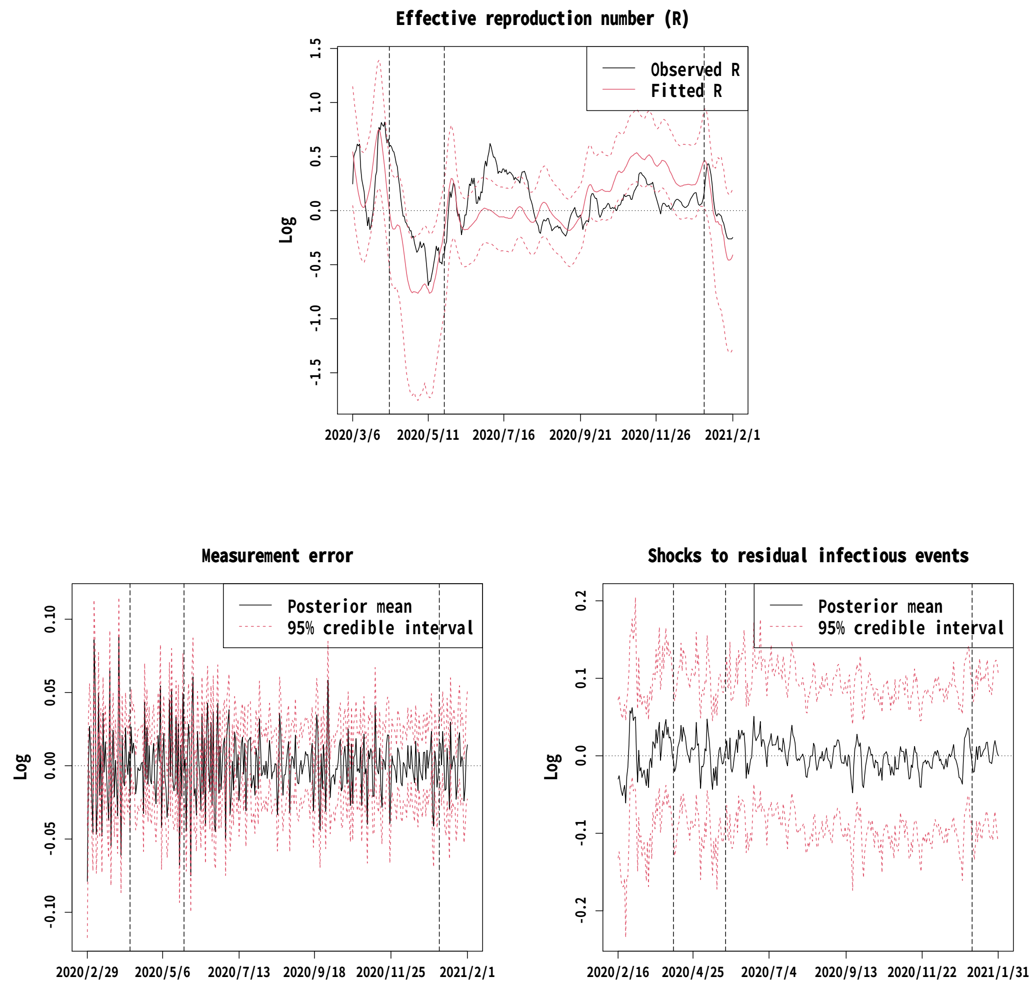
Source: Sugishita, Kurita, Sugawara, and Ohkusa (2020).

Figure 6: Dummy variable for absolute humidity



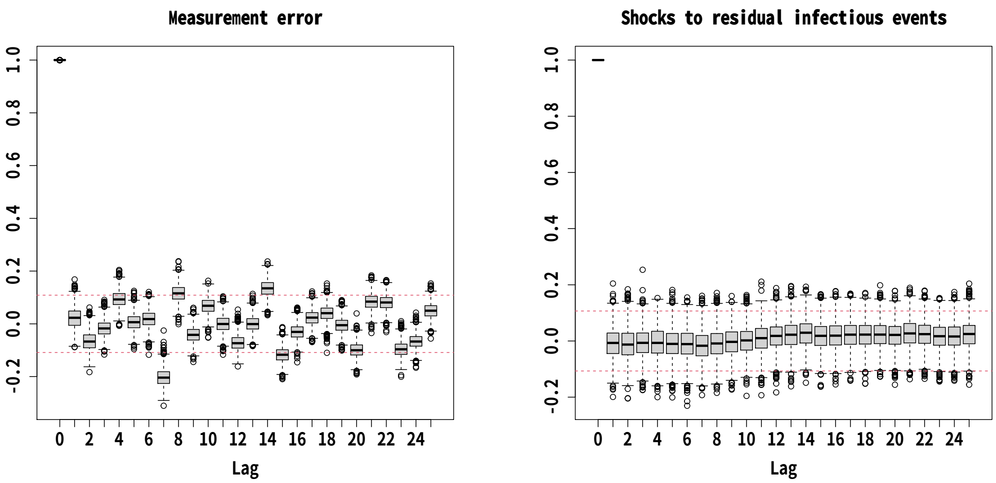
Notes: The figure plots the daily value of $D_{AH,t}$. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021, to June 20, 2021.

Figure 7: Fitted value of the effective reproduction number and residuals



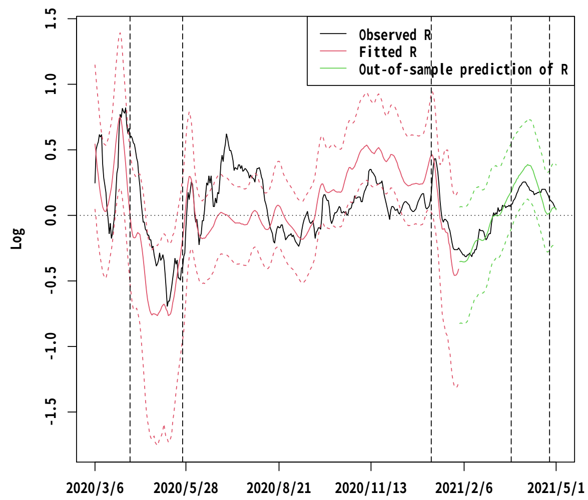
Notes: In the top panel, "Observed R" indicates the log of the observed daily estimate of the effective reproduction number; and "Fitted R" indicates the fitted value of the log of the daily estimate of the effective reproduction number in the regression model estimated by 2020-21 data. Red dashed lines in each panel indicate the 95% credible interval. In the bottom panels, "Measurement error" and "Shocks to residual infectious events" indicate the values of η_t and ϵ_t, respectively. In both top and bottom panels, vertical dashed lines are the first and the last dates of two states of emergency: from April 7, 2020, to May 25, 2020; and from January 7, 2021, to March 21, 2021.

Figure 8: Mcmc samples of auto-correlation functions of residuals



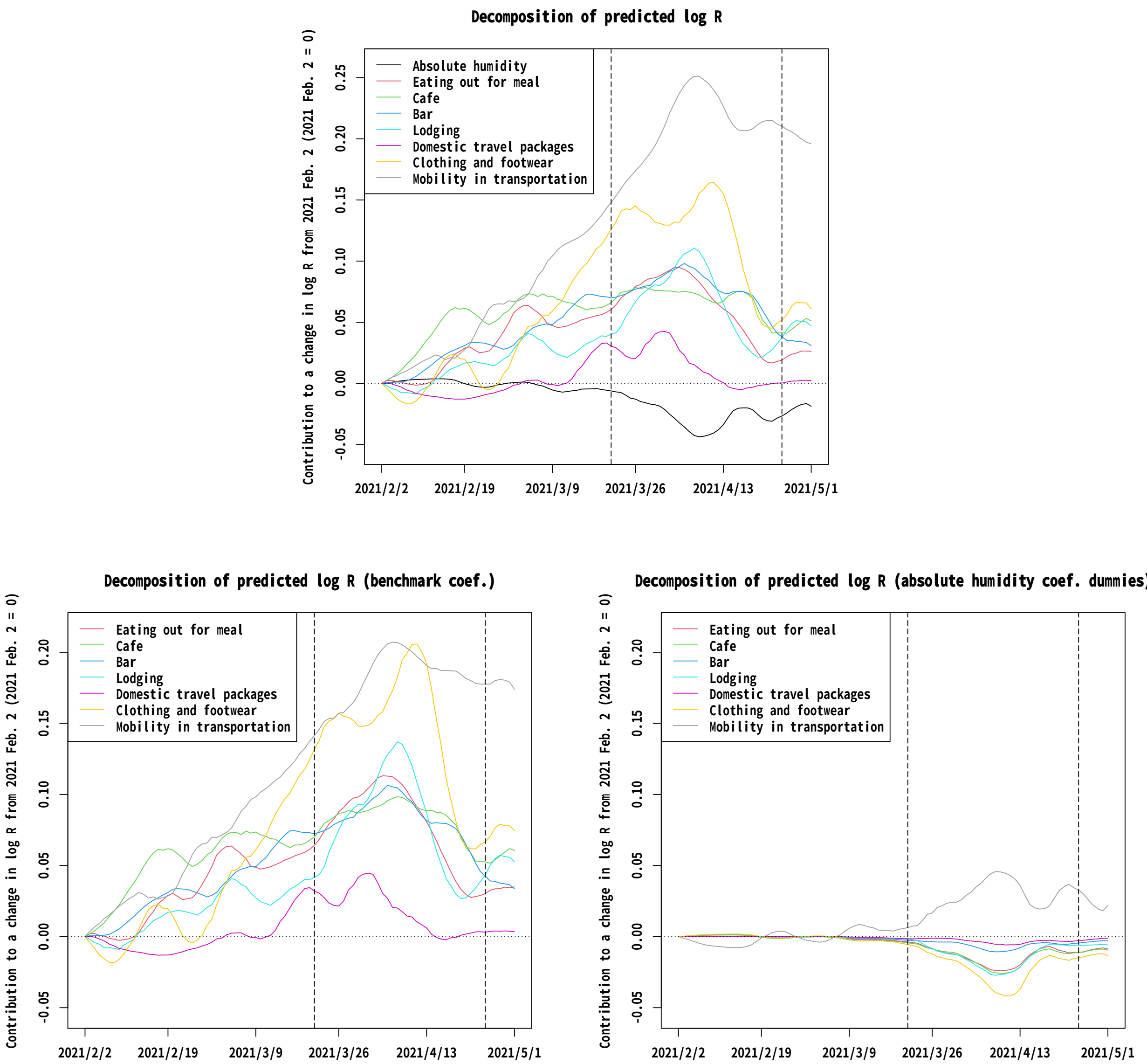
Notes: “Measurement error” and “Shocks to residual infectious events” indicate η_t and ϵ_t, respectively. For each lag, the grey box shows the range between 25% and 75% percentiles, and the black line in the middle of the box indicates the median. The whiskers extended above and below the box show the range between 25% percentile - 1.5*(75% percentile-25% percentile) and 75% percentile + 1.5*(75% percentile-25% percentile). Each circle indicates the value of an outlier outside this range.

Figure 9: Out-of-sample prediction of the effective reproduction number from February 2, 2021, to May 1, 2021



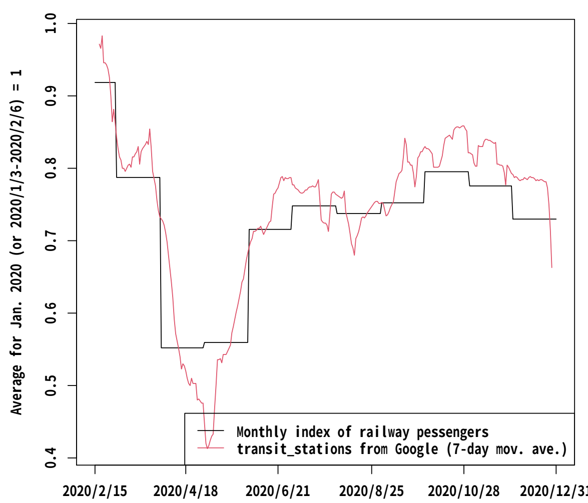
Notes: "Observed R" indicates the log of the observed daily estimate of the effective reproduction number; "Fitted R" indicates the fitted value of the log of the daily estimate of the effective reproduction number in the regression model; and "Out-of-sample prediction of R" indicates the predicted value of the log of the daily estimate of the effective reproduction number from February 2, 2021, to May 1, 2021, generated by the regression model estimated with data up to February 1, 2021. Red and green dashed lines indicate the 95% credible intervals.

Figure 10: Factor decomposition of the effective reproduction number from February 2, 2021, to May 1, 2021



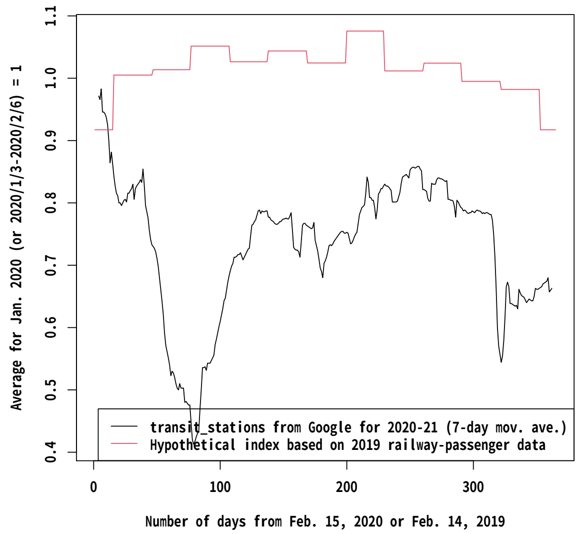
Notes: The top panel shows changes from February 2, 2021, in the sum of the product of each explanatory variable and the posterior mean of its linear coefficient, and the product of the cross term between each explanatory variable and the absolute humidity dummy and the posterior mean of its coefficient. The bottom-left panel shows only the first product, and the bottom-right panel shows only the second product, for each explanatory variable.

Figure 11: The number of railway passengers and mobility in public transportation in 2020



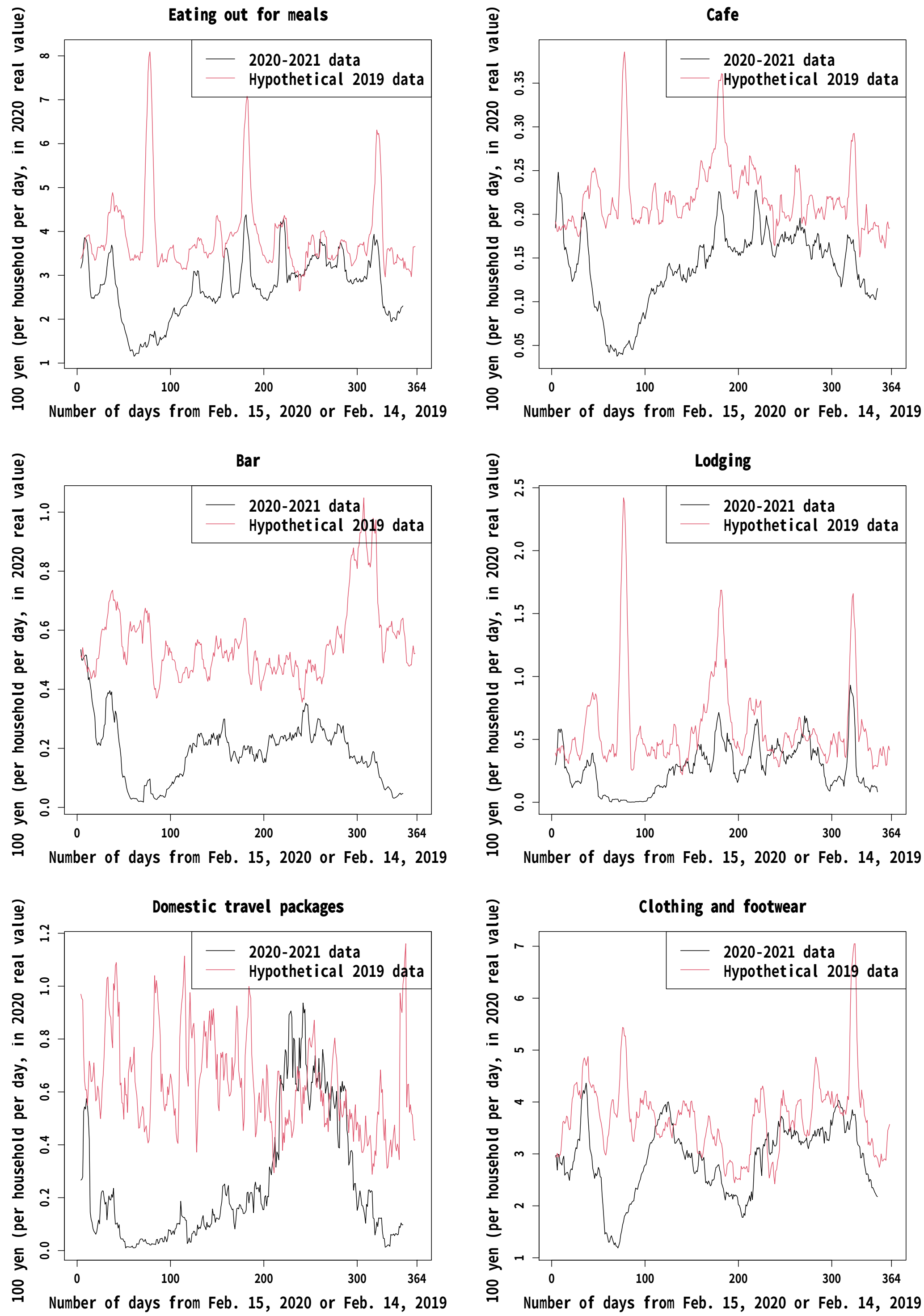
Notes: "transit_stations" is a measure of mobility in public transportation in the COVID-19 Community Mobility Reports from Google, which is available from February 15, 2020. For this measure, a 7-day centered moving average is shown in the figure. The index of railway passengers is constructed by dividing the monthly average of railway passengers in each month of 2020 by the monthly average in January 2020. The monthly value of this index is shown for each date within the same month.

Figure 12: Mobility in public transportation in 2019 and for 2020-21



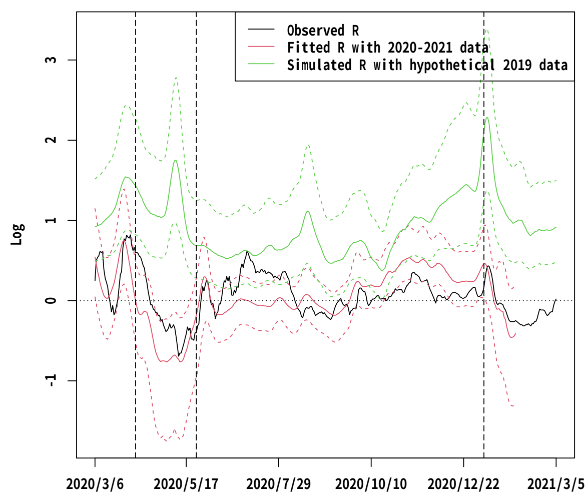
Notes: "transit_stations" is a measure of mobility in public transportation in the COVID-19 Community Mobility Reports from Google, which is available from February 15, 2020. The figure for this measure is a 7-day centered moving average. The index of railway passengers is constructed by dividing the monthly average of railway passengers in each month of 2019 by the monthly average in January 2020. The monthly value of this index is shown for each date within the same month. The index starts from February 14, 2019, and then is connected with its value on January 1, 2019, after the year end, so that it loops as a hypothetical index of mobility in public transportation without any restriction on household consumption or mobility.

Figure 13: Real household expenditures per household in 2019 and for 2020-21



Notes: In each panel, the 2019 data start from February 14, 2019, and are connected with the data on January 1, 2019, after the year end, so that they loop for 365 days as hypothetical data without any restriction on household consumption or mobility. The 2020-21 data start from February 15, 2020, and end at January 31, 2021. All figures are 7-day centered moving averages.

Figure 14: Simulated effective reproduction number without any restriction on household consumption or mobility



Notes: The vertical axis is the log of the daily estimate of the effective reproduction number. "Observed R" is the log of the observed daily estimate of the effective reproduction number. "Fitted R with 2020-2021 data" is the fitted value of the log of the daily estimate of the effective reproduction number in the regression model estimated by 2020-2021 data. "Simulated R with hypothetical 2019 data" is the log of the daily estimate of the effective reproduction number simulated by the regression model with hypothetical values of explanatory variables based on 2019 data. Vertical dashed lines are the first and the last dates of two states of emergency: from April 7, 2020, to May 25, 2020; and from January 7, 2021, to March 21, 2021.