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

Hajime Tomura*

June 17, 2021

Abstract

Daily estimates of the effective reproduction number for new coronavirus based on reporting dates are regressed on real household expenditures per household for eating out, traveling, and apparel shopping, as well as mobility in public transportation, using publicly available nationwide data in Japan from February 15, 2020, to February 1, 2021. The effects of absolute humidity, the declaration of states of emergency, and the year-end and new-year holiday period are controlled through dummy variables. The lagged infectious effects of explanatory variables due to incubation periods are also taken into account. Out-of-sample prediction of the estimated regression model traces closely the realized values of the effective reproduction number from February 2 to May 1 in 2021. The factor decomposition of the fourth wave of infection in April 2021 indicates that increases in mobility in public transportation and household expenditure for apparel shopping had the largest infectious effects among the explanatory variables, separately from eating out and traveling. Estimated regression coefficients indicate that real household expenditures for cafe and bar had larger effects on the effective reproduction number per value of spending than the other types of household expenditures in the explanatory variables during the sample period. Thus, a loss of aggregate demand will be minimized if the effective reproduction number is lowered by restricting household consumption of cafe and bar.

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 Waseda University for their valuable comments. First draft: April 30, 2021.

1 Introduction

Household consumption activities have been regarded as part of main causes that spread new coronavirus infection by generating human contacts. To quantify this causal relationship, I regress 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 nationwide data in Japan. These real household expenditures are included in the explanatory variables because they have been the main subjects of government interventions, or have shown a high sample correlation with the effective reproduction number. Because the definition of the effective reproduction number is the number of new cases of infection per an infected person in the current population, real household expenditures in the explanatory variables are also normalized on a per-household basis. In this paper, I use nationwide data because there is no publicly available household expenditure data for each prefecture at daily frequency. Because of data availability and the spread of mutant strains in 2021, the latter of which may cause a structural change in the regression model, the sample period of data used for the estimation of the regression is set to the period between February 15, 2020, and February 1, 2021.

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 date. Then, the log of the effective reproduction number on each date is modeled as 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 coefficients to explanatory variables through cross terms between explanatory variables and time dummies. The regression model incorporates a white noise as measurement error of the effective reproduction number, and also a latent AR(1) process for unobserved 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 effective reproduction number 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 the predicted values of the effective reproduction number trace the realized values closely. This result indicates that correlation between the effective reproduction number and the explanatory variables had been stable up to April 2021, and also that a bias in the regression model is small.

Given the successful out-of-sample prediction, I decompose a surge in the effective reproduction number 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 effect on the surge in the effective reproduction number in April 2021; household expenditure for clothing and footwear had the second largest effect; and household expenditures for eating out for meal, cafe, bar, and lodging had effects of similar magnitudes. This result implies that, even after controlling for the effects of infectious household expenditures, an increase in mobility in public transportation had a significant infectious effect. It also implies that there was an infectious effect of apparel shopping separately from eating out at bars and restaurants, which tend to happen after shopping on the high street.

In addition, estimated regression coefficients imply that real household expenditures for cafe and bar had larger effects on the effective reproduction number per value of spending than the other household expenditures included in the regression model during the sample period. Thus, a loss of aggregate demand will be minimized if the effective reproduction number is lowered by restricting household consumption of cafe and bar. Given this estimation result, I run counterfactural simulations to quantify the effect of restricting cafe and bar consumption on the effective reproduction number, using the estimated regression model. The simulations imply that it will be necessary to cut more than 80% of cafe and bar consumption by households compared to the 2019 level in order to keep the annual average of the effective reproduction number below one, unless the infectiousness of cafe and bar consumption is reduced.

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 marginal economic 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, self-restricting behavior, and social interventions, this paper takes a bottom-up approach, providing reduced-form estimates of the effects of detailed household expenditures and mobility 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

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

3. The estimation of the regression model and the out-of-sample prediction of the estimated regression model are reported in section 4. Counterfactual simulations using the estimated regression model are described in section 5. 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 effective reproduction number on a selected number of household expenditure items 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 these variables, and explain the reasons for the selection of explanatory variables in the regression.

2.1 Data sources

Table 1 summarizes the sources of data used in this paper. The effective reproduction number published by Toyokeizai-Shinpo-Sha, a publisher in Japan, is the week-over-week gross rate of change in the 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 that it takes for an infected person to cause the next infection) and 7 is the number of days in the reporting interval, which is a week. This simplified formula to estimate the effective reproduction

[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.

number based on reporting dates has been widely used in Japan to update the effective reproduction number real time daily.[3] In the Family Income and Expenditure Survey, daily data on nominal household expenditures are publicly available only for households with two or more members.

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

Figure 1 plots 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 effective reproduction number being an exponential function of the week-over-week gross rate of change in the number of new cases. The sample period starts from March 1, 2020, as the effective reproduction number from the data source is published only from that date.

I focus on these household expenditure items, because the first five items have been the main subjects of government interventions into household consumption. 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, and from January 7 to March 21 in 2021, and prohibited the sales of alcohol at bars and restaurants in metropolitan area during the third state of emergency from April 25, 2021. Also, the government subsidized domestic traveling for sightseeing 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-

[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]Nominal household expenditure for foreign travel packages was negligible during the sample period.

To-Travel" campaign. There is a controversy over whether this campaign helped spreading new-coronavirus infection across the country.

Regarding clothing and footwear, this item has been showing a high sample correlation with the effective reproduction number, as shown in Figure 2. In fact, clothing and footwear has a higher maximum cross correlation coefficient with the effective reproduction number than any other large category of nominal household expenditures, and as high a maximum cross correlation coefficient as nominal household expenditure for bar (see Table 2).

For these reasons, I will consider the six household expenditure items shown in Figure 1 for explanatory variables in the regression. Given the limited length of the sample period, I do not include other household expenditure items in the regression, in order to limit the number of explanatory variables.

2.3 Sample correlation between the effective reproduction number and mobility

Figure 3 plots 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 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 multicollinearity 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 will be used as a general measure of mobility. This choice is due to convenience, as it allows me to use publicly available transportation data in 2019 for a substitute to Google data when I simulate the estimated regression model with data for the period before mobility data from Google are available, as will be described later.

3 Regression model

3.1 Regression model and the definition of variables

Given the discussion described in the previous section, I regress the log 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 definition of the effective reproduction number is the number of new cases per an infected person in the current population, real household expenditures in the explanatory variables are also normalized on a per-household basis.

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 of the proportional subsidies for domestic travels during the "Go-To-Travel" campaign. 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.[5]

[5]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 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 as part of 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 costs, nominal household expenditure for domestic travel packages increased substantially during the campaign period, while that for lodging did not. To remove the effect of the Go-To-Travel campaign from the CPI for lodging, I linearly interpolate the monthly CPI for lodging between July 2020 and January 2021 when converting nominal household expenditure for lodging in real terms. There is a corresponding CPI for clothing and footwear.

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-i} and η_{t-i} in the past 7 days, including the current date (i.e., for $i = 0, 1, ..., 6$), because the log of the effective reproduction number on the left-hand side is equivalent to the sum of the rate of change in the number of new cases in the past 7 days, multiplied by 5/7.

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, 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 daily infectious events k days ago, i.e., V_{t-k}, for $k = 1, 2, ..., 14$, to measure the contribution from infectious events in k days ago for the rate of change in the 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 expenditure items per household and mobility in public transportation, which are denoted by $X_{i,t}$ for $i = 1, 2, ..., 7$. 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 $i = 1, 2, ..., 7$ spread new-coronavirus infection to some extent.

To compute $D_{AH,t}$ for each date, the dummy for absolute humidity no less than $9g/m^3$ for 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 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.[6] The end of the sample period is set to include explanatory variables only up to January 2021 in the regression. 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.[7] By February 10, 2021, 108 cases of mutant strains had been found nationwide.[8] Also, the Tokyo Metropolitan 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.[9] Thus, the spread of mutant strains was likely to be limited before the end of January 2021.

[6] 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).

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

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

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

3.3 Possible biases in the regression model

Before moving on, let me clarify possible biases in the regression model, as the model is in a linear reduced form with a limited number of explanatory variables. 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 significantly, 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. Therefore, time-specific infectious events during the other holiday periods, such as the Golden Week, are included in unobserved infectious events, e_t, in the regression model. This set-up may violate the assumption that e_t follows the same AR(1) process throughout the sample period.

There is no immediate simultaneous equation bias in the regression model, because the dependent variable is the daily sum of the rates of increase in reported new cases of infection over the past 7 days, and all the explanatory variables for the rate of increase in reported new cases on each day are lagged variables due to incubation periods, as implied by (2). However, if household expenditures hold perfect foresight or rational expectations of future effective reproduction numbers, it is possible to consider a case that violates the assumption that explanatory variables are uncorrelated with the error term (i.e., η_t) in the regression model.[10] 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 unobserved infectious activities, e_t, include some household activities that are correlated with the explanatory variables in the regression model. Given the difficulty to resolve all the concerns on possible biases due to a small sample size and limited data availability, I will instead compare out-of-sample prediction of the estimated regression model with observed data to see if a bias in the regression model is small in the next section.[11]

4 Estimation result

I apply the Bayesian method to estimate parameters in the regression. I set an uninformative, or improper, prior distribution for each parameter, that is, the density of the prior

[10]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}:

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

where γ, θ, and ϕ are constant and ν_t is an independent white noise. If households have perfect foresight, $E_{t-1}R_t = R_t$; thus, 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.

[11]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 dependent variable and explanatory variables through this channel 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.

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.[12]

Table 4 shows the posterior mean and the 95% credible interval of each parameter value. The fitted value of the log of the effective reproduction number and also the residuals of the regression are shown in Figure 7. The fitted value deviates from the observed 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 unobserved 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 unobserved infectious events, (4), in the regression.[13]

4.1 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 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 of this paper.[14] Note that the prediction period still includes the fourth wave of infection in April 2021 in Japan,

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

[13]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.

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

during which there was a surge in the effective reproduction number across Japan.

Figure 9 compares the predicted and realized values 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 correlation between the effective reproduction number and the explanatory variables in the regression model had been stable up to April 2021. It also confirms that the regression model does not have a large bias, despite several concerns on specification error described in section 3.3.

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 state of emergency on March 21, 2021, then the predicted values 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 of the estimated regression model, Figure 10 decomposes changes in the predicted values of the effective reproduction number from the trough in February 2021 to the peak in April 2021 into contributions of the explanatory variables in the regression model. The top panel shows the total effects from both the linear coefficient to each explanatory variable and the coefficient to the cross term between the absolute humidity dummy and each 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 effect on the surge in the effective reproduction number during April 2021; household expenditure for clothing and footwear had the second largest effect; household expenditures for eating out for meal, cafe, bar, and lodging had effects of similar magnitudes; and household expenditure for packaged domestic travels had a small effect.

This result implies that, even after controlling for the effects of infectious household expenditures, an increase in mobility in public transportation had a significant infectious effect. It also implies that there was an infectious effect of apparel shopping separately from eating out at bars and restaurants, which tend to happen after shopping on the high street. Even though the identification of the infectious activities behind this result is beyond the scope of this paper, a possible explanation is that droplets due to oral conversations in congested coaches, stations, and apparel shops spread new coronavirus infection. There may be overlooked infectious activities specific to apparel shopping as well, such as fitting.

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 a minor impact on the result described above.[15]

5 Counterfactual simulations

5.1 Subject of simulations

Table 4 shows that 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. 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 cafe and

[15] The contribution of mobility in public transportation in the top-right panel of Figure 10 is positive, because both the coefficient to the cross term with the absolute humidity dummy and the measure of mobility in public transportation are negative. Thus, an increase in absolute humidity reduces the degree of a reduction in the effective reproduction number due to a given decline in mobility in public transportation from the benchmark period (i.e., 2020 January 3rd to February 6th).

bar consumption by households.[16]

5.2 Standard for policy evaluation

Hereafter, I simulate the quantitative effect of restricting cafe and bar consumption on the effective reproduction number, using the estimated coefficients of the regression model. For the measure to evaluate policy effects, I use the geometric mean of simulated effective reproduction numbers over a year. I highlight this indicator because if the effective reproduction number remains above one on average, then the number of new cases will exceed the finite capacity of medical services at some point in the future. Even though choosing a year for the duration of the simulation period implies a pessimistic expectation that vaccinations will be widely available in the country only after a year, it allows to take into account the seasonality of household expenditures fully in the simulations.[17]

5.3 Benchmark simulation with hypothetical 2019 data

To set a benchmark, I first simulate 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. 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. This indexation is consistent with the feature of the COVID-19 Community Mobility Reports such that each type of mobility data in the reports are

[16]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 in the third state of emergency since April 25, 2021.

[17]To clarify, the government may face a trade-off between new cases of new-coronavirus infection (or deaths) and a measure of household activities such as GDP, if it aims to stabilize the effective reproduction number at some specific level between 0 and 1 until the arrival of vaccinations for a sufficiently large part of the population, because the total number of deaths due to new coronavirus will be lower as the targeted value of the effective reproduction number is set closer to zero. This question is beyond the scope of this paper.

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 effective reproduction number for 365 days from March 6, which coincides with the first date of the effective reproduction number in the regression model with 2020-2021 data.[18] To simulate 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 values 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 rise 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.[19]

[18]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.

[19]A caveat is that the Golden Week holiday period in 2019 lasted for 10 days, which was longer than usual.

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 the previous section indicating that a bias in the regression model is small.

5.4 Quantitative effect of restricting cafe and bar consumption by households on the effective reproduction number

Hereafter, I simulate the quantitative effects of percentage reductions of cafe and bar consumption by households compared to the 2019 level. In the simulation, I keep the values of the other household expenditures unchanged from the 2020-2021 data, except domestic travel packages. This is because there has been a self-restraint on packaged domestic travels in 2020, except for the Go-To-Travel campaign period between July 22 and December 27 in 2020, as shown in Figure 13. To take into account this observation, it is assumed that real household expenditure per household for domestic travel packages will be as low as the average in the period between the end of the first state of emergency and the beginning of the Go-To-Travel campaign period, i.e., from May 26 to July 21 in 2020. Also, I do not use the realized values of mobility in public transportation up to the end of the first state of emergency, because there was an one-off adjustment in mobility from the pre-pandemic level

Thus, an increase in household consumption during the Golden Week in 2019 could be higher than that in a regular year. I thank Hiroshi Fujiki for pointing out this anomaly in 2019.

to the pandemic level, and then an one-off large drop in mobility due to the declaration of the first state of emergency during this period. Instead, up to the end of the first state of emergency on May 25, 2020, the daily values of mobility in public transportation are set to the average of transit_stations in the COVID-19 Community Mobility Reports from Google from June to November in 2020. I still use the realized daily values of transit_stations for the subsequent period to take into account seasonal fluctuations in mobility, such as those during weekends and holidays. See Figure 15 for the hypothetical series of real household expenditure per household for domestic travel packages and mobility in public transportation assumed in the simulation. I set zero to all dummies related to the states of emergency (i.e., $D_{SE,j,t} = 0$ for $j = 0, 1, 2$), while keeping 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., $D_{NY,t}$).

Table 6 shows the posterior distribution of annual means of $\ln R_t$ in the simulation. The table implies that the government cannot keep the annual geometric average of the effective reproduction number below 1, unless it cuts 95-100% of cafe and bar consumption by households compared to the 2019 level. It can be shown that even if the response of mobility in public transportation to a restriction on cafe and bar consumption is endogenized in the simulation, the government would need to cut 80-85% of cafe and bar consumption to keep the annual geometric average of the effective reproduction number below 1. See the 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 contacts, I regress the log of the estimate of the effective reproduction number based on reporting dates 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 estimated regression model closely traces the observed effective reproduction number from the trough in February 2021 to the fourth wave of infection in April 2021. This result implies that correlation between the effective reproduction number and the explanatory variables in the regression model had been stable up to April 2021, and that a bias in the regression model is small. The factor decomposition of out-of-sample prediction indicates that increases in mobility in public transportation and household expenditures for clothing and footwear made the largest contributions to the fourth wave of infection in April 2021 among the explanatory variables, separately from eating out and traveling. Thus, it is important to identify whether there are overlooked infectious activities in public transportation and apparel shopping.

The estimation result indicates that a loss of aggregate demand will be minimized if the effective reproduction number is lowered by cutting household consumption of cafe and bar. Counterfactual simulations using the estimated regression model, however, indicate that it would be necessary to cut more than 80% of cafe and bar consumption compared to the 2019 level, in order to keep the annual geometric average of the effective reproduction number below one.

A caveat to this result is that I ignore the substitution and complementarity among household expenditure items when I simulate the effect of a restriction on cafe and bar consumption by households. Also, it would be more realistic to consider state-dependent interventions by the government in response to seasonal fluctuations in household consumption and mobility, and hence the effective reproduction number. Due to the challenge to counterfactual simulation analysis described above, however, the simulations of more detailed intervention policies are beyond the scope of this paper. With these reservations, counterfactual simulations in this paper provide ballpark estimates of the effect of a cost-effective intervention in household consumption on the effective reproduction number.

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 Simulating the effect of a restriction on cafe and bar consumption by households with an endogenous response of mobility in public transportation

To endogenize the response of mobility in public transportation to a restriction on household consumption, I regress transit_stations in the COVID-19 Community Mobility Reports from Google (i.e., $X_{7,t}$) on real household expenditures per household (i.e., $X_{i,t}$ for $i = 1, 2, ..., 6$) among the explanatory variables of the regression model for the effective reproduction number. I also include the following time dummies as part of explanatory variables to capture the seasonality of mobility: holidays, including weekends and the year-end and new-year holiday period from December 29-January 3; each of the two states of emergency from April 7, 2020, to May 25, 2020 and from January 7, 2021, to March 21, 2021; and December, which is due to a change in the relationship between bar consumption and mobility due to year-end parties. I estimate the regression coefficients by OLS. The sample period is from February 15, 2020, to January 31, 2021, which is the same as the sample period of explanatory variables in the estimation of the regression model for the effective reproduction number.

Among the explanatory variables, eating out for meals (i.e., $X_{1,t}$) has a statistically insignificant coefficient. Table 7 reports the OLS estimate of the regression of transit_stations without eating out for meals in the explanatory variables.

To see the fit of this regression by an out-of-sample prediction, Figure 16 plots the ratio of the monthly average of railway passengers in 2019 to the January 2020 average, and the monthly average of daily fitted values generated by applying estimated regression coefficients shown in Table 7 to the 2019 data of the explanatory variables. To compute the fitted values, time dummies for the two states of emergency are set to zero. The difference between the means of the two series in the figure can be interpreted as a time fixed effect. The figure shows that the fitted values largely replicate the observed pattern of time variations in the number of railway passengers in 2019, such that the number of railway passengers drops significantly in February and December, while fluctuating around a stable level from March to November. Hereafter, I use the regression coefficients shown in Table 7 when I endogenize the response of mobility in public transportation to an exogenous restriction on household consumption. In so doing, time dummies for the two states of emergency remains to be set to zero.

Table 8 shows the posterior distributions of annual means of $\ln R_t$ when mobility in public transportation is endogenized by the regression coefficients shown in Table 7. The table implies that the government would need to cut 80-85% of cafe and bar consumption by households compared to the 2019 level, in order to stabilize the effective reproduction number below 1 on average throughout a year.

Table 1: Data sources

Data	Level	Frequency	Source
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
Railway passengers	Nationwide	Monthly	Statistical Survey on Railway Transport, Ministry of Land, Infrastructure, Transport and Tourism
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 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 effective reproduction number is available only from March 1, 2020.

Table 3: Definition of variables

R_t	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	Unobserved infectious events.
ϵ_t	Shocks to unobserved infectious events.
η_t	Measurement error.

Notes: The effective reproduction number is the week-over-week gross rate of change in the 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, the 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 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 regression model for the effective reproduction number. 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.

Table 6: Posterior distribution of annual means of $\ln R_t$ with restrictions on cafe and bar consumption

Degree of % reduction of cafe and bar consumption compared to the 2019 level	Posterior mean	2.5% percentile	97.5% percentile
50%	0.30	0.00	0.63
55%	0.26	-0.04	0.60
60%	0.23	-0.09	0.56
65%	0.20	-0.13	0.53
70%	0.17	-0.17	0.51
75%	0.14	-0.22	0.49
80%	0.11	-0.26	0.47
85%	0.08	-0.31	0.45
90%	0.05	-0.37	0.44
95%	0.02	-0.42	0.42
100%	-0.00	-0.48	0.40

Notes: Each figure is the annual mean of $\ln R_t$ for an exogenous percentage reduction of cafe and bar consumption by households compared to the 2019 level in the first column. For each figure, 0 corresponds to the case in which the geometric annual mean of the effective reproduction number is 1.

Table 7: Regression of mobility in public transportation on real household expenditures per household

	OLS estimate	Standard deviation	t value
Intercept	0.653	0.014	45.66
$X_{2,t}$	0.608	0.089	6.82
$X_{3,t}$	0.211	0.026	7.90
$X_{4,t}$	-0.045	0.014	-3.12
$X_{5,t}$	0.038	0.010	3.75
$X_{6,t}$	0.011	0.004	2.66
Dummy for holidays	-0.147	0.01	-12.97
Dummy for the first state of emergency	-0.141	0.013	-10.51
Dummy for the second state of emergency	-0.063	0.014	-4.54
Dummy for December	0.025	0.025	1.00
$X_{3,t}$*(Dummy for December)	-0.179	0.138	-1.29

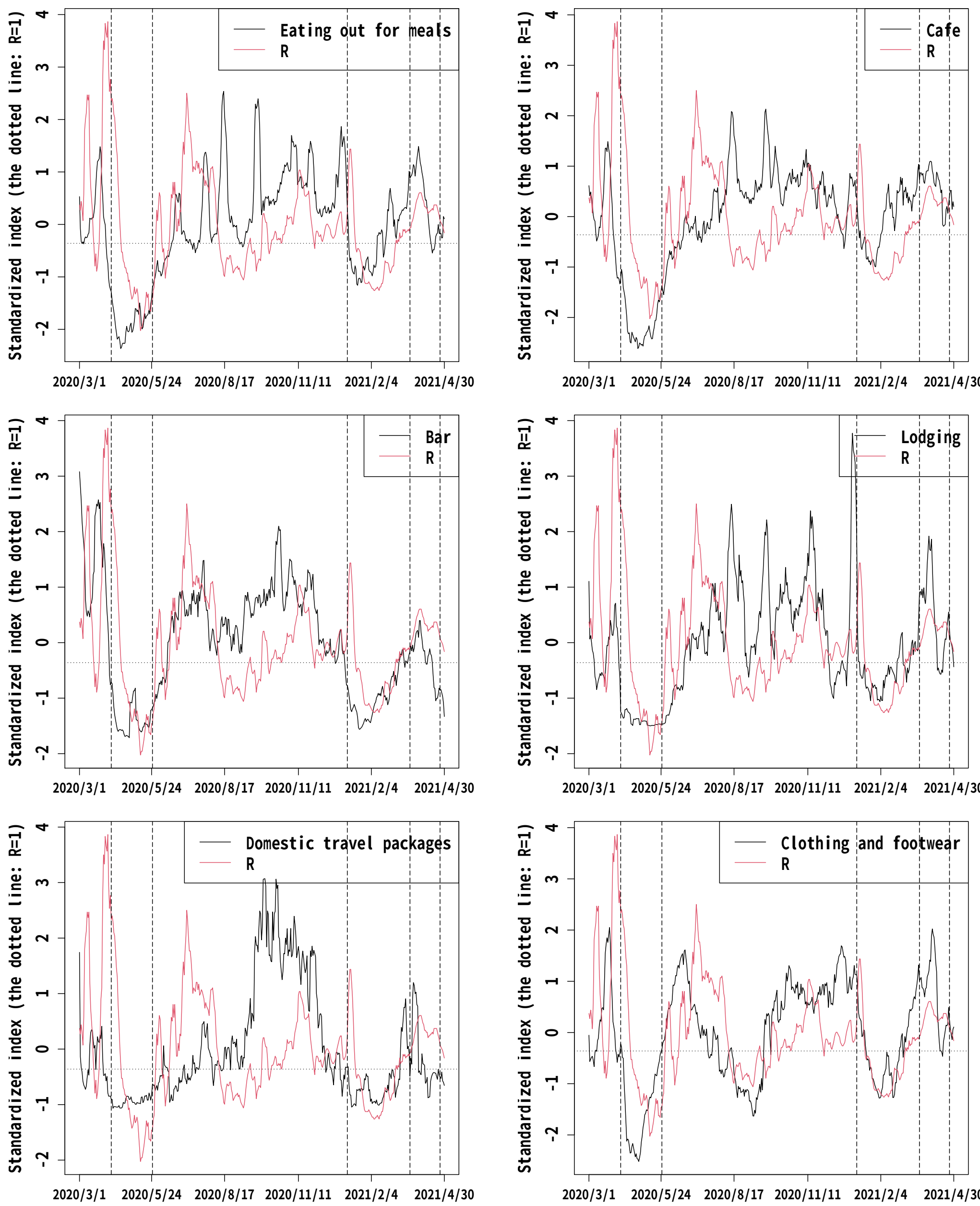
Dependent variable: $1 + X_{7,t}/100$.
R^2: 0.77; adj. R^2: 0.76.
Sample period: Februrary 15, 2020, - January 31, 2021.

Table 8: Posterior distribution of annual means of $\ln R_t$ with restrictions on cafe and bar consumption and endogenous responses of mobility in public transportation

Degree of % reduction of cafe and bar consumption compared to the 2019 level	Posterior mean	2.5% percentile	97.5% percentile
50%	0.29	-0.00	0.62
55%	0.24	-0.06	0.58
60%	0.20	-0.11	0.53
65%	0.16	-0.16	0.49
70%	0.12	-0.22	0.46
75%	0.08	-0.29	0.42
80%	0.03	-0.35	0.39
85%	-0.00	-0.41	0.36
90%	-0.04	-0.48	0.34
95%	-0.08	-0.55	0.31
100%	-0.13	-0.63	0.28

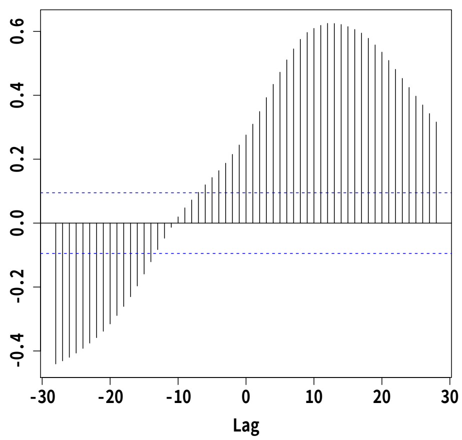
Notes: In this simulation, endogenous responses of mobility in public transportation are taken into account. Each figure is the annual mean of $\ln R_t$ for an exogenous percentage reduction of cafe and bar consumption by households compared to the 2019 level in the first column. For each figure, 0 corresponds to the case in which the geometric annual mean of the effective reproduction number is 1.

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



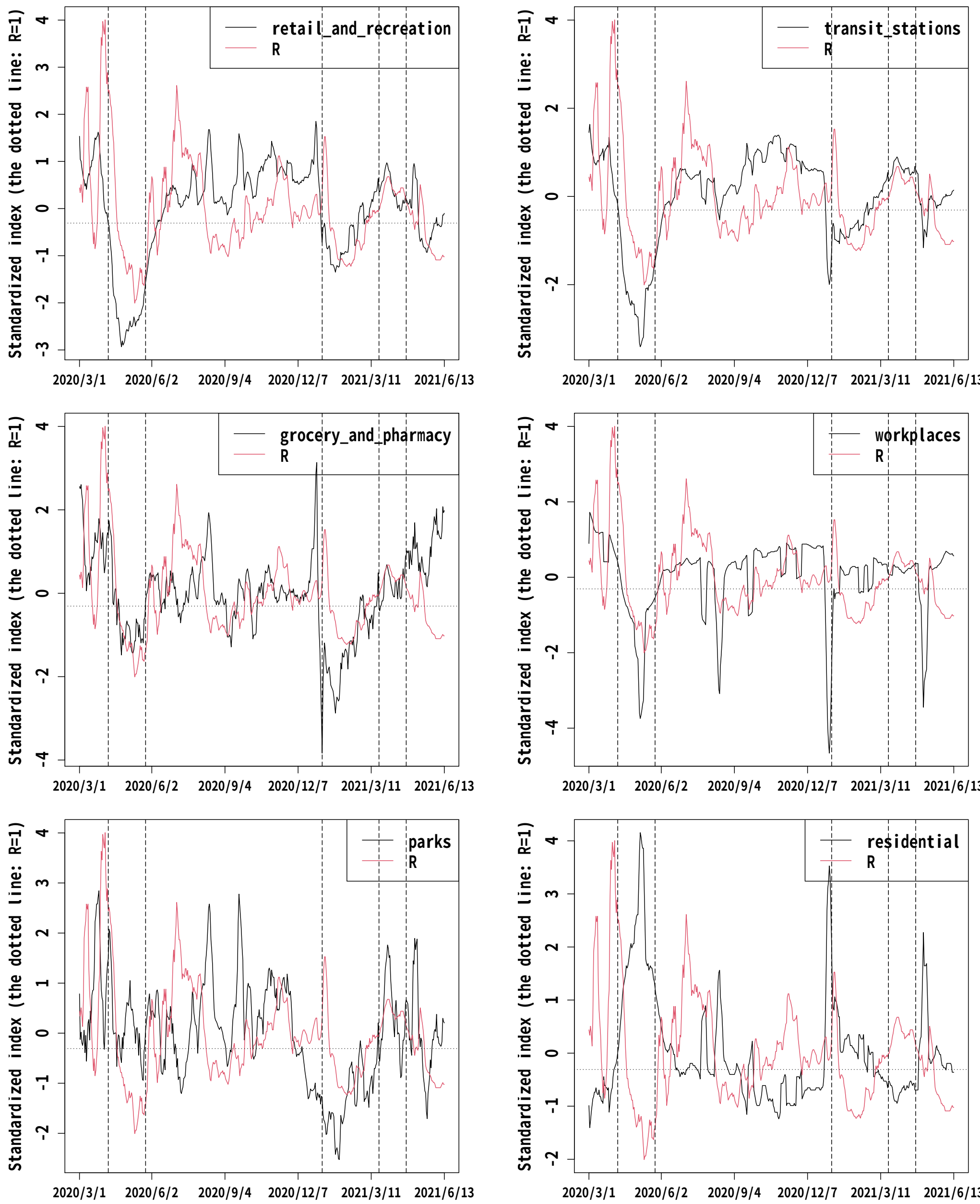
Notes: In each panel, "R" indicates the effective reproduction number, and nominal household expenditure per household is a 7-day backward moving average. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021. 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



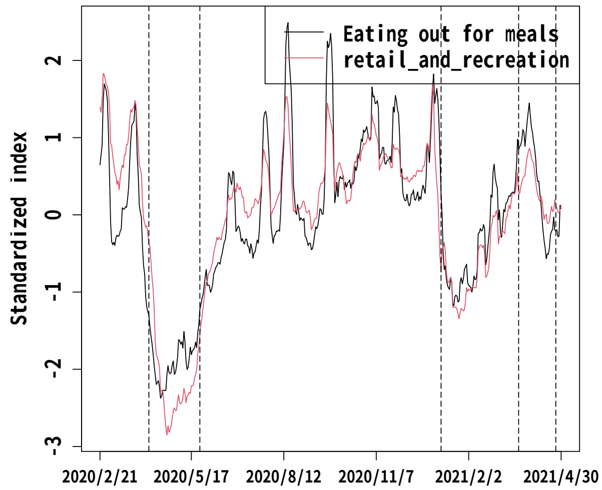
Notes: The figure shows the correlation coefficient between the contemporaneous effective reproduction number and lagged 7-day backward moving averages of nominal household expenditure per household for clothing and footwear. On the horizontal axis, negative lags are leads. Horizontal dashed lines are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

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



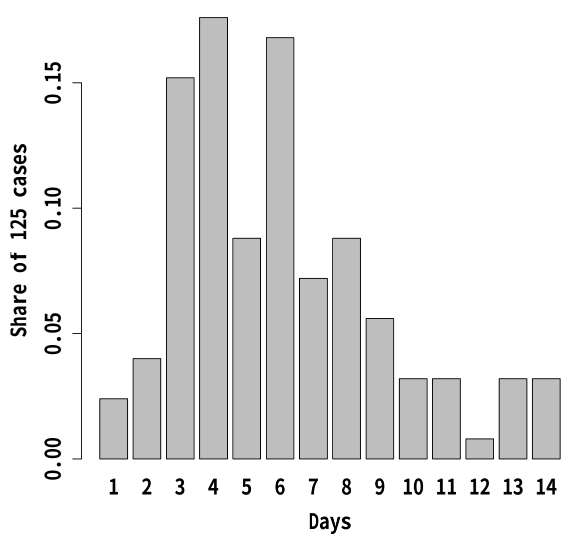
Notes: In each panel, "R" is the effective reproduction number, and the measure of mobility is a 7-day backward moving average. All figures are standardized by their means and standard deviations. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021.

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



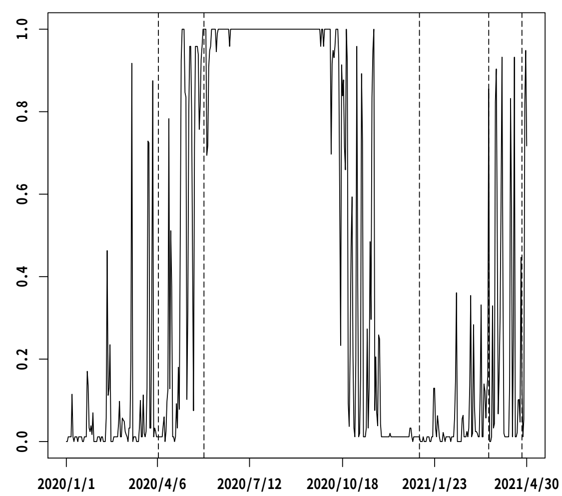
Notes: The figure plots retail_and_recreation in the COVID-19 Community Mobility Reports from Google and real household expenditure per household on eating out for meals. Both figures are 7-day backward moving averages, and standardized by their means and standard deviations. Vertical dashed lines are 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.

Figure 5: A sample distribution of incubation periods in Japan



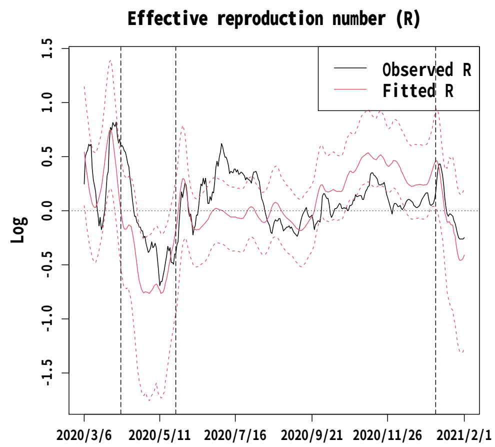
Source: Sugishita, Kurita, Sugawara, and Ohkusa (2020).

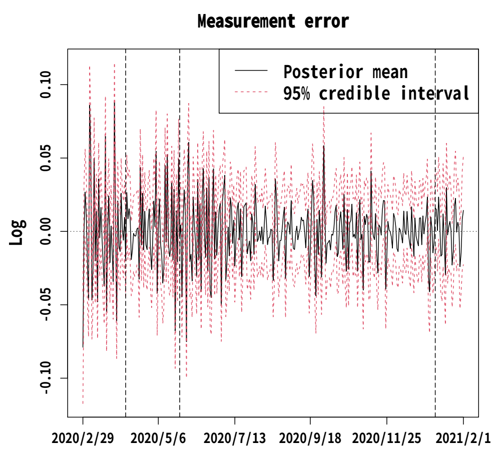
Figure 6: Dummy variable for absolute humidity

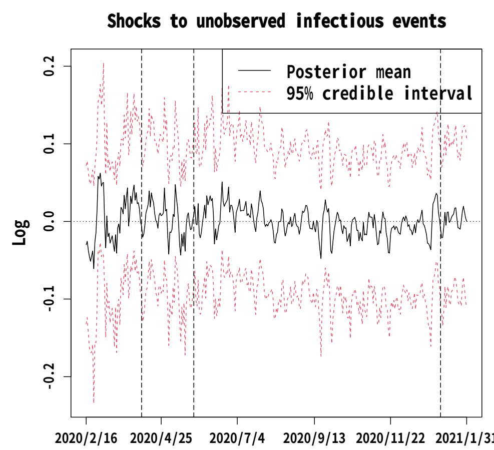


Notes: The figure plots the daily value of $D_{AH,t}$. Vertical dashed lines are the first and the last dates of three states of emergency: from April 7, 2020, to May 25, 2020; from January 7, 2021, to March 21, 2021; and from April 25, 2021.

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

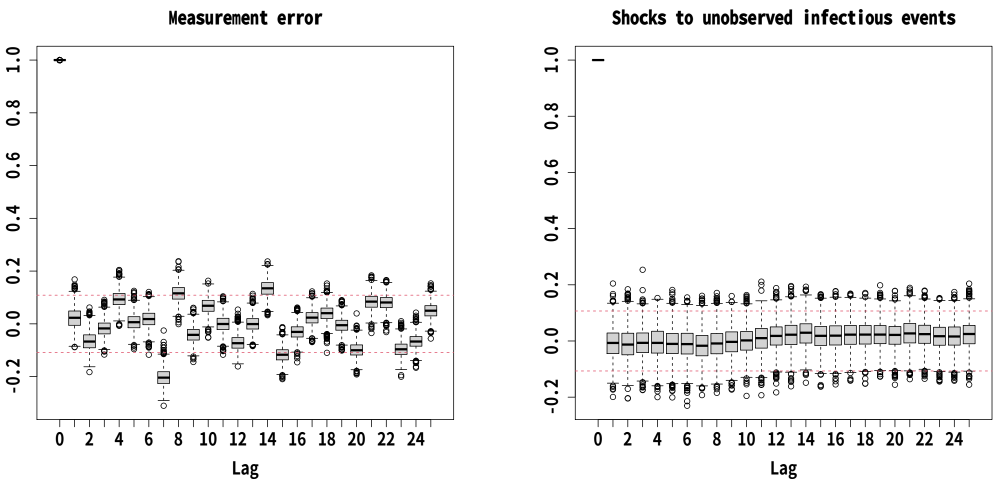






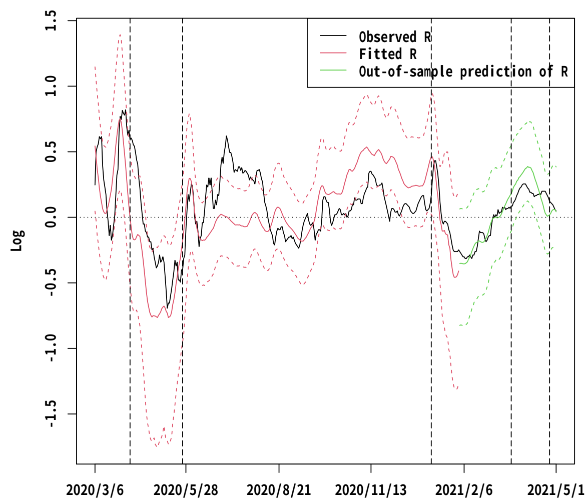
Notes: In the top panel, "Observed R" indicates the log of the observed effective reproduction number; and "Fitted R" indicates the fitted value of the log of the effective reproduction number in the regression model with 2020-21 data. Red dashed lines in each panel indicate the 95% credible interval. In the bottom panels, "Measurement error" and "Shocks to unobserved 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



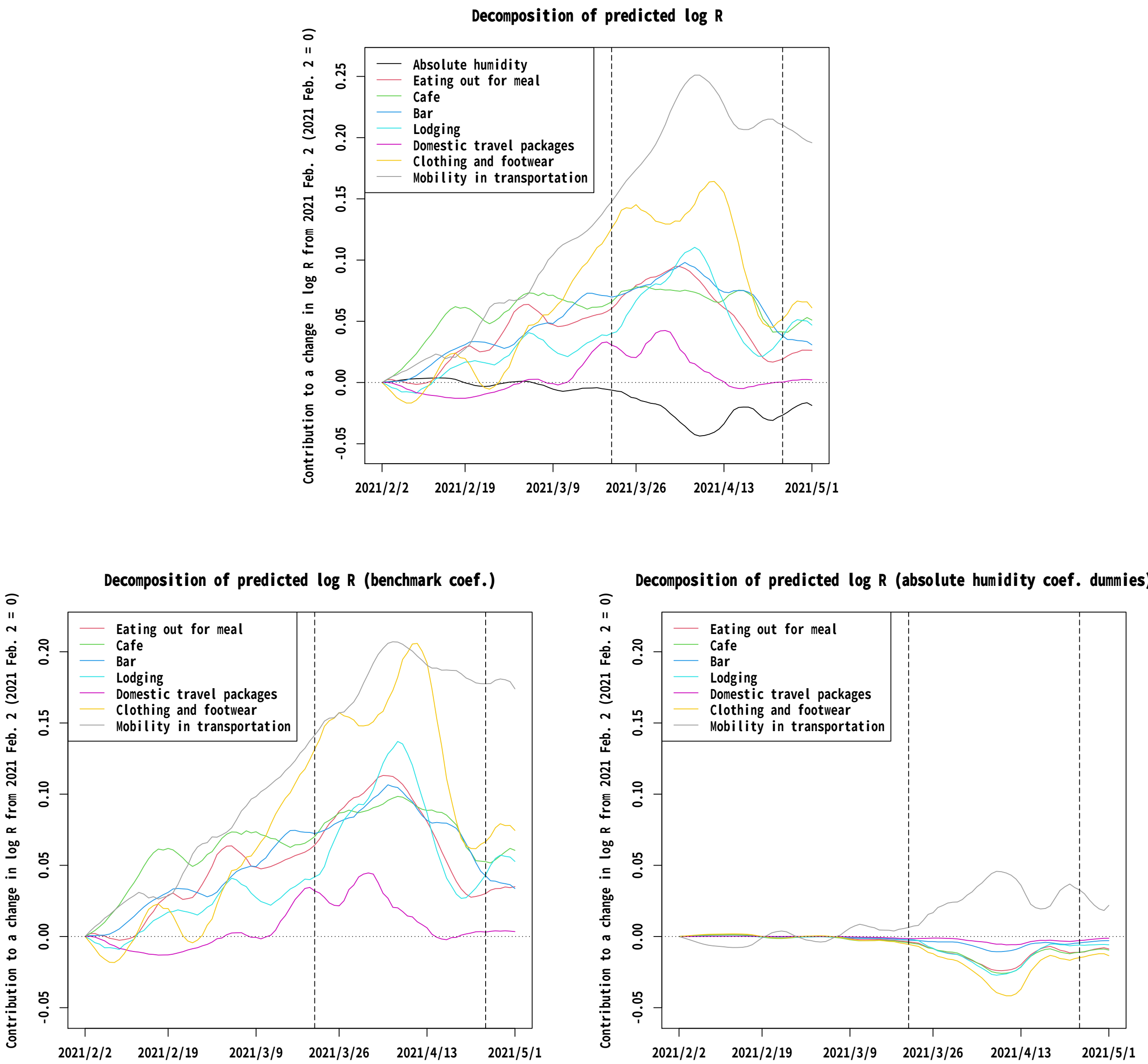
Notes: "Measurement error" and "Shocks to unobserved infectious events" indicate η_t and ϵ_t, respectively. For each lag, the grey box shows the range between 25% and 75% percentiles, and the black line in the middle of the box indicates the median. The whiskers extended above and below the box show 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



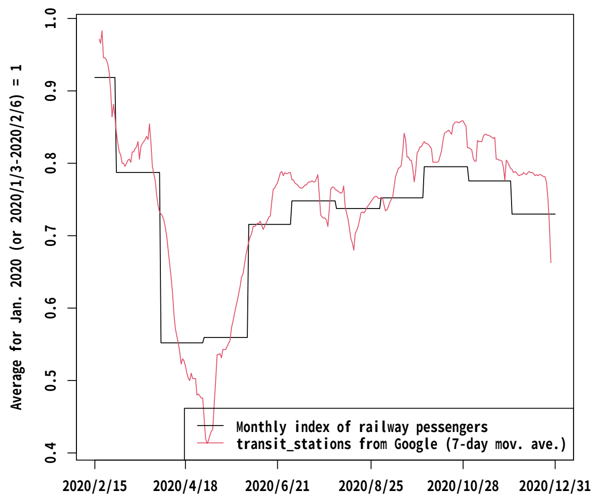
Notes: "Observed R" indicates the log of the observed effective reproduction number; "Fitted R" indicates the fitted value of the log of the effective reproduction number in the regression model estimated with data up to February 1, 2021; and "Out-of-sample prediction of R" indicates the predicted value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

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



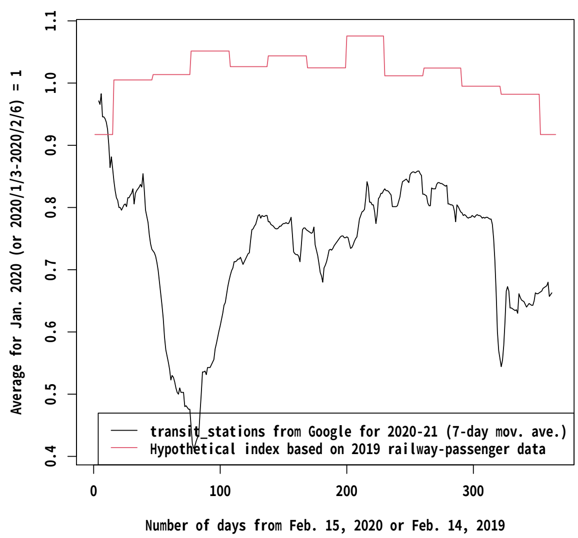
Notes: The top panel shows changes from February 2, 2021, in the product of each explanatory variable and the posterior mean of its estimated coefficient in the regression model, including the cross term with the absolute humidity dummy. The bottom-left panel excludes the cross term between each explanatory variable and the absolute humidity dummy from the figures in the top panel, and the bottom-right panel extracts only changes associated with the cross term between each explanatory variable and the absolute humidity dummy.

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



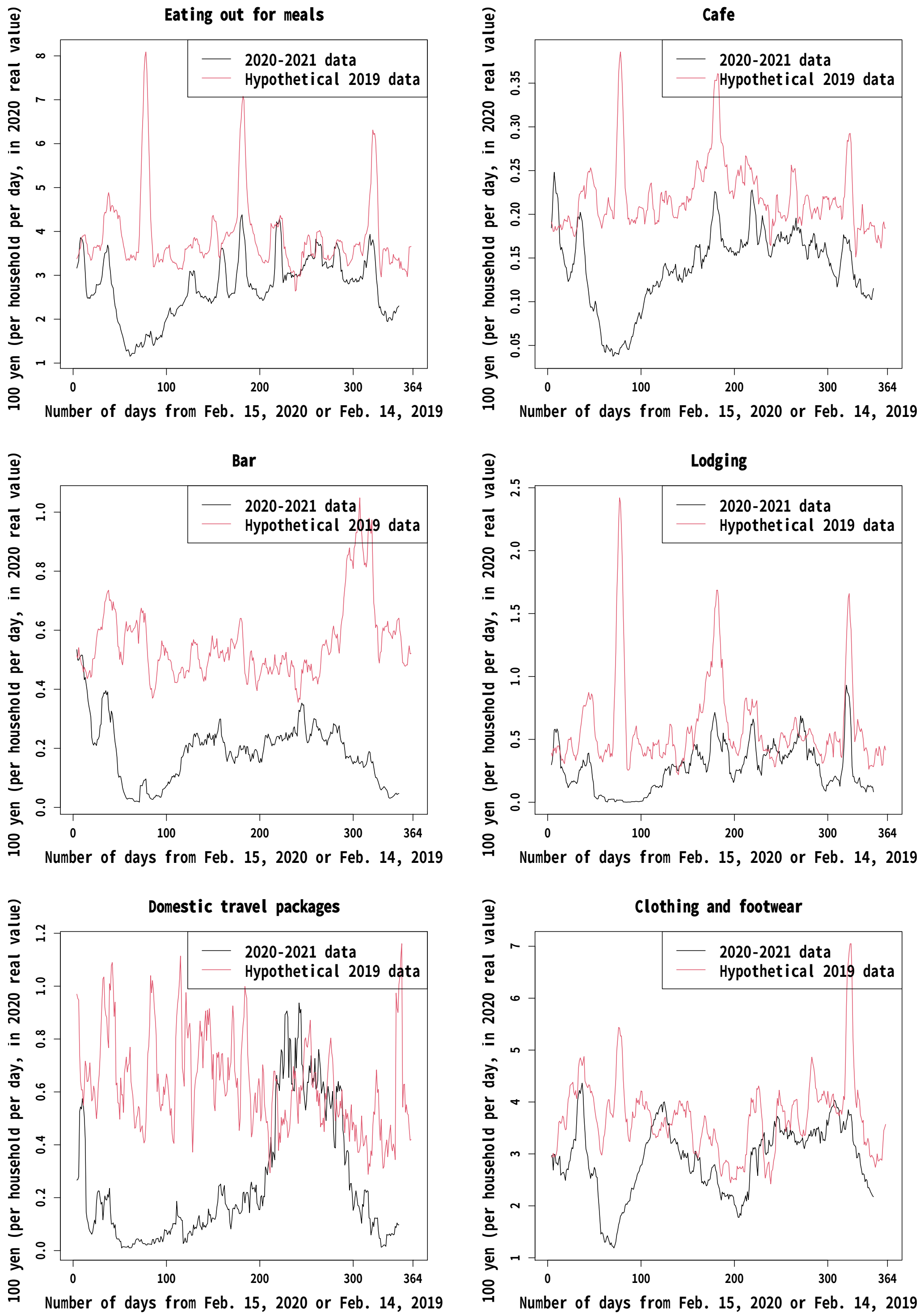
Notes: "transit_stations" is a measure of mobility in public transportation in the COVID-19 Community Mobility Reports from Google, which is available from February 15, 2020. For this measure, a 7-day centered moving average is shown in the figure. The index of railway passengers is constructed by dividing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



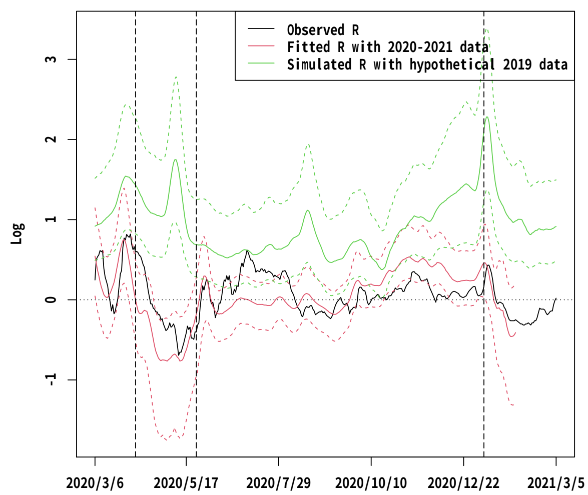
Notes: "transit_stations" is a measure of mobility in public transportation in the COVID-19 Community Mobility Reports from Google, which is available from February 15, 2020. The figure for this measure is a 7-day centered moving average. The index of railway passengers is constructed by dividing the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



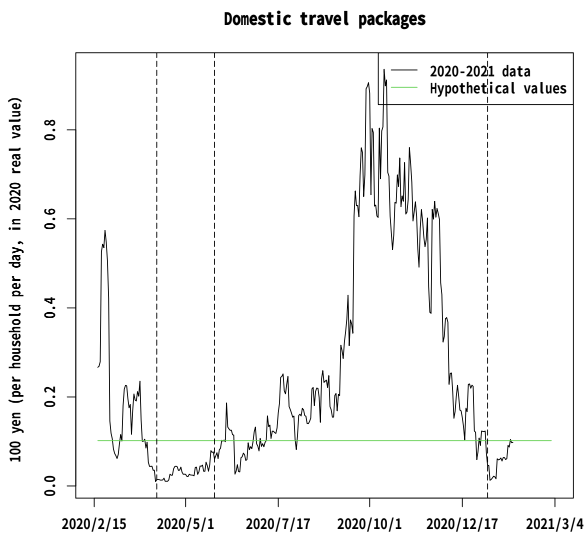
Notes: In each panel, the 2019 data start from February 14, 2019, and are connected with the data on January 1, 2019, after the year end, so that they loop for 365 days as hypothetical data without any restriction on household consumption or mobility. The 2020-21 data start from February 15, 2020, and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

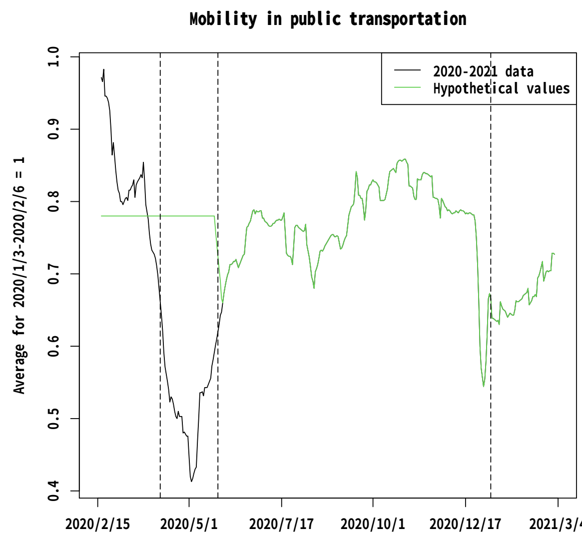
Figure 14: Simulated effective reproduction number without any restriction on household consumption or mobility



Notes: The vertical axis is the log of the effective reproduction number. “Observed R” is the log of the observed effective reproduction number. “Fitted R with 2020-2021 data” is the fitted value of the log of the effective reproduction number in the regression model with 2020-2021 data. “Simulated R with hypothetical 2019 data” is the daily value of $\ln R_t$ simulated by the regression model for the effective reproduction number 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.

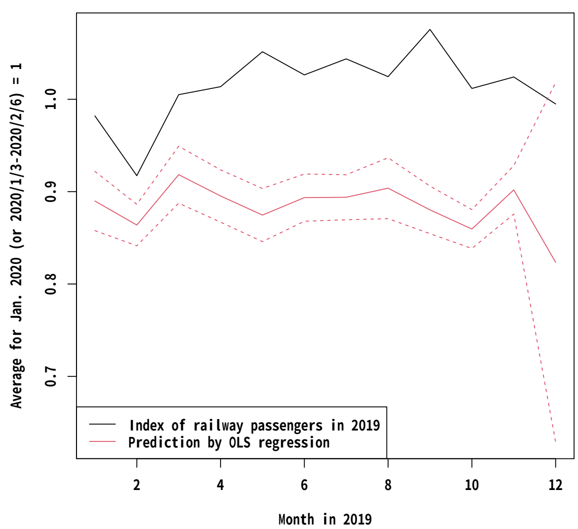
Figure 15: Hypothetical real household expenditure per household for domestic travel packages and mobility in public transportation for the simulation of restrictions on cafe and bar consumption





Notes: In each panel, "2020-2021 data" are the observed data for 2020-2021, and "Hypothetical values" are values for the simulation. Up to the end of the first state of emergency on 2020 May 25, the hypothetical values of mobility in public transportation are set to the average of transit_stations in the COVID-19 Community Mobility Reports from Google during June-November in 2020; and they are set to the realized daily values of transit_stations after 2020 May 25. The hypothetical values of real household expenditure per household for domestic travel packages are set to the average in the period between the end of the first state of emergency and the beginning of the Go-To-Travel campaign period, i.e., from May 26, 2020, to July 21, 2020. All series in the figure are 7-day centered moving averages.

Figure 16: The observed and fitted value of mobility in public transportation in 2019



Notes: “Index of railway passengers in 2019” is the ratio of the monthly average of railway passengers in each month of 2019 to the January 2020 average. ”Prediction by OLS regression” is the fitted value generated by inserting 2019 data in the explanatory variables of the regression shown in Table 7, except that time dummies for the two states of emergency are set to zero. The daily fitted values are averaged out to compute the monthly average for each month in the figure. Dotted lines around ”Prediction by OLS regression” indicate the 95% confidence interval.