LOCKDOWNS AS PARTIAL HOLIDAYS: IDENTIFYING WORK-STOPPAGE ELASTICITIES FROM RECURRING HOLIDAYS AND TRANSFERRING THEM TO FORECAST THAILAND'S MONTHLY EXPORTS THROUGH THE COVID-19 SHOCK

A PREPRINT

Thanatchon Chongmankhong

November 7, 2026

ABSTRACT

The COVID-19 pandemic produced a collapse-and-rebound episode in Thailand's monthly export series that contaminates the trend and seasonal components learned by forecasting models and, more fundamentally, cannot be anticipated by any parameter estimated before the shock. This paper proposes the *Holiday-Equivalent Stoppage* (HES) framework, which treats a lockdown as a partial holiday. Recurring holidays—Thai public holidays including Songkran (domestic channel) and the Chinese New Year window (partner channel)—identify a work-stoppage elasticity through one *shared* regressor per channel in a decomposable forecasting model (Prophet). The elasticity is transferred to the pandemic through a Holiday-Equivalent Stoppage Index (HESI) that expresses each lockdown day in holiday units, computed as the drop in workplace mobility relative to its drop on actual holidays. We derive (i) an error decomposition showing that the transfer is exact when the dose is correctly scaled and that calibration absorbs the projection of unmodelled shocks on the dose path; (ii) a Jensen-type aggregation bias under which monthly models attenuate a daily elasticity β to approximately $e^{\beta} - 1$; (iii) the resulting identification problem with monthly data together with a closed-form aggregation correction; and (iv) dose saturation in tree ensembles. A Monte Carlo study with 30 replications, in which export values are simulated while the holiday calendar and the OxCGRT and Google Mobility data are real, confirms each prediction. With a correctly scaled dose, daily Prophet reduces the 2020–2021 mean absolute percentage error (MAPE) from 26.0% to 7.8%, whereas switching holiday effects on for whole months (22.7%) or using raw stringency (27.2%) does not help. Daily estimation recovers the elasticity (-0.80 against a true -0.80), monthly estimation yields -0.54, close to the predicted $e^{\beta} - 1 = -0.55$, and LightGBM saturates. After the shock, SARIMAX and simply masking the shock window remain competitive. HES is therefore a structured tool for measurable stoppage shocks rather than a universal accuracy improvement.

Keywords Forecasting under structural shocks · Calendar effects · Prophet · COVID-19 · Temporal aggregation · Tree ensembles · International trade · Thailand

JEL codes C15, C22, C53, F17

1 Introduction

Thailand is among the most trade-dependent economies in Asia, and its monthly merchandise-export statistics are one of the most closely watched indicators of the business cycle. They are forecast routinely by government agencies, the central bank and private institutions, typically with univariate or calendar-augmented time-series models and, increasingly, with machine-learning methods. The COVID-19 pandemic confronted all of these models with an episode unlike anything in their estimation samples: a sharp contraction in the second quarter of 2020, followed by a rebound

in 2021 that was driven at least as much by global demand as by the lifting of domestic restrictions (Baldwin and Tomiura, 2020). Episodes of this kind are a principal source of forecast failure (Clements and Hendry, 1998; Rossi, 2013).

Such episodes create three distinct problems. The first is *contamination*: models that learn trend and seasonal components from the data absorb the shock into these components, so forecasts made after the shock inherit its distortions. The second is *propagation*: lagged features carry the anomaly forward, so a collapse in 2020 depresses forecasts for 2021 through year-on-year lags. The third, and most fundamental, is *unforecastability*: a model has no parameter describing the effect of a lockdown until a lockdown has been observed. The first two problems concern estimation after a shock; the third concerns forecasting during one.

The literature offers many tools for the first two problems: intervention analysis (Box and Tiao, 1975), joint estimation of outlier effects (Chen and Liu, 1993), structural-break estimation (Bai and Perron, 1998), estimation-window selection (Pesaran and Timmermann, 2007), volatility rescaling of vector autoregressions after March 2020 (Lenza and Primiceri, 2022), and analogue forecasting based on earlier recessions (Foroni et al., 2022). Practitioners using Prophet are advised to encode lockdown periods as one-off holidays (Meta, n.d.). All of these remedies identify the effect of a shock *after* observing it. Tree ensembles, which dominate recent forecasting competitions (Makridakis et al., 2022), face an additional obstacle: their prediction functions are piecewise constant and cannot extrapolate beyond the support of the training data (Hastie et al., 2009; Januschowski et al., 2022).

This paper starts from a simple observation. From the point of view of production and shipping, a lockdown resembles a holiday: factories, offices, ports and customs operate at reduced capacity because people do not go to work. Thailand runs a natural experiment of this kind every year. Songkran, the Thai New Year in mid-April, and the other public holidays close large parts of the economy for one to five days, and the Chinese New Year closes Thailand's largest trading partner for about a week. These events recur in every year of a pre-pandemic sample, so their effect on exports is identified long before the pandemic. If the intensity of a lockdown could be expressed in holiday units—"a lockdown day is equivalent to 0.3 of a holiday"—the holiday elasticity would deliver a forecast of the lockdown effect without any new parameter.

The naive implementation of this idea, switching the holiday effect on for every day of a lockdown month, implicitly assumes that a lockdown day is a full holiday, which is rarely true. We therefore replace the switch with a *dose*: the Holiday-Equivalent Stoppage Index (HESI), defined as the drop in workplace mobility on a given day divided by the typical drop on a public holiday, both measured in the Google Community Mobility Reports (Google LLC, 2022). The holiday serves as the unit of measurement that converts mobility into stoppage. We address five research questions:

RQ1 *Transfer.* Can an elasticity identified from recurring holidays forecast exports during lockdowns that never occur in the estimation sample?

RQ2 *Dose.* How sensitive is the transfer to the scale of the dose, and can the scale be calibrated in real time?

RQ3 *Frequency.* What is lost when only monthly data are available, and can it be recovered?

RQ4 *Model class.* Can tree ensembles perform the same transfer?

RQ5 *Aftermath.* How should the shock be treated when forecasting after it has ended?

The paper makes four contributions. First, it formulates the HES framework, in which one shared regressor per channel carries both recurring holidays and pandemic doses, and it makes the identifying assumptions explicit. Second, it derives the theoretical properties of the framework: an error decomposition with an explicit optimal dose; a Jensen-type aggregation bias with the closed form $e^{\beta} - 1$; a closed-form aggregation correction with its delta-method variance; the identification problem of monthly data; the dose saturation of tree ensembles; and the reason why one-off holidays cannot transfer. Third, it evaluates the framework in a Monte Carlo study designed according to the ADEMP structure of Morris et al. (2019), built on the real Thai holiday calendar and on real policy and mobility data, with paired inference across replications. The theory predicts the simulation outcomes quantitatively, not only qualitatively. Fourth, it provides an open implementation and a protocol for official export statistics.

The main findings are as follows. With a correctly scaled dose, transfer works: daily Prophet reduces the 2020–2021 MAPE from 26.0% to 7.8% without observing a single pandemic day, and it improves on the no-dose forecast in 100% of replications. The whole-month switch and raw stringency do not help. When the mobility dose overstates the true stoppage two-fold, calibrating the scale on the first two lockdown months lowers the average MAPE from 13.0% to 8.8% and significantly improves on the uncalibrated dose, although the gain over the no-dose forecast is not consistent enough across replications to survive a multiple-testing correction. Monthly estimation attenuates the elasticity almost exactly as predicted, and the aggregation correction removes the bias at the cost of a variance that the delta method also predicts. Tree ensembles switch rather than dose. After the shock, a simple SARIMAX model and masking the shock window are at least as accurate as HES, and in normal times tree ensembles are the most accurate method.

All quantitative results in this paper come from a simulation study; applying the framework to official Thai export statistics, which the accompanying code supports, is left for future work.

Section 2 reviews related work. Section 3 presents the framework and Section 4 its theoretical properties. Section 5 describes the simulation design and Section 6 the results. Section 7 discusses implications, threats to validity and a protocol for official data, and Section 8 concludes. Proofs are collected in Appendix A.

2 Related literature

Calendar and moving-holiday effects. The modelling of trading-day and holiday effects in monthly series has a long tradition in official statistics. Cleveland and Devlin (1982) and Bell and Hillmer (1983) formulate regression models for calendar variation, and moving-holiday regressors for festivals such as Easter became a standard component of X-12-ARIMA (Findley et al., 1998). For Asian economies, the Lunar New Year shifts between January and February and induces large month-to-month swings that require dedicated regressors (Lin and Liu, 2003). This literature estimates holiday effects in order to remove them; we estimate them in order to *reuse* them as a source of identifying variation for a different shock.

Structural change, interventions and forecast failure. Intervention analysis (Box and Tiao, 1975) and the joint estimation of outlier effects (Chen and Liu, 1993) model known or detected disturbances with dummy-type regressors whose coefficients are estimated from the disturbance itself. Tests and estimators for structural breaks (Perron, 1989; Bai and Perron, 1998), evidence of pervasive instability (Stock and Watson, 1996) and the theory of forecast failure (Clements and Hendry, 1998; Rossi, 2013) explain why such disturbances degrade forecasts, and estimation-window selection (Pesaran and Timmermann, 2007) offers one response. In all of these approaches the size of the disturbance is learned from the disturbance; none provides a forecast of its effect before it is observed.

Forecasting through the pandemic. Lenza and Primiceri (2022) show how to estimate vector autoregressions after March 2020 by rescaling shock volatility, and Foroni et al. (2022) use the dynamics of the Global Financial Crisis to forecast the COVID-19 recession and recovery. Both rely on similarity of *magnitude* with earlier events. Our approach relies on similarity of *mechanism* with events that recur every year. On the trade side, Baldwin and Tomiura (2020) emphasise that the pandemic hit trade through supply disruptions, demand collapse and supply-chain contagion; HES targets the first channel and treats the others as unmodelled shocks, a distinction that matters for calibration (Proposition 1).

Decomposable models and Prophet. Prophet (Taylor and Letham, 2018) represents a series as the sum of a piecewise-linear trend, Fourier seasonality, holiday effects and extra regressors, estimated by maximum a posteriori under Gaussian and Laplace priors; see also Hyndman and Athanasopoulos (2021) for a textbook treatment of related decompositions. Each holiday name receives its own coefficients, which is convenient for seasonal holidays but implies that a new "COVID" holiday has no information before it occurs (Remark 1).

Tree ensembles. Regression trees (Breiman et al., 1984) and gradient boosting (Friedman, 2001), in particular LightGBM (Ke et al., 2017), are among the most accurate methods for tabular data (Grinsztajn et al., 2022) and in large forecasting competitions (Makridakis et al., 2022). Their prediction functions are piecewise constant and flat outside the training support (Hastie et al., 2009; Januschowski et al., 2022). We show that this property prevents them from transferring a dose even when the dose variable is supplied explicitly.

Temporal aggregation and mixed frequency. The effects of temporal aggregation on time-series models are surveyed by Silvestrini and Veredas (2008); temporal disaggregation (Chow and Lin, 1971; Denton, 1971) and mixed-data-sampling regressions (Ghysels et al., 2007) address the mismatch between high- and low-frequency information. We add a specific result: under a log link, holiday effects concentrated on a few days and lockdown effects spread over a month aggregate differently, which biases monthly estimates of a daily elasticity in a predictable way.

Policy and mobility data. The Oxford COVID-19 Government Response Tracker (Hale et al., 2021) provides daily stringency indices, and the Google Community Mobility Reports (Google LLC, 2022) provide daily changes in visits to workplaces and other places relative to a pre-pandemic baseline. Mobility data have been used to nowcast economic activity (Sampi and Jooste, 2020). We use them in a different role: as a unit converter between lockdown intensity and holiday-equivalent stoppage.

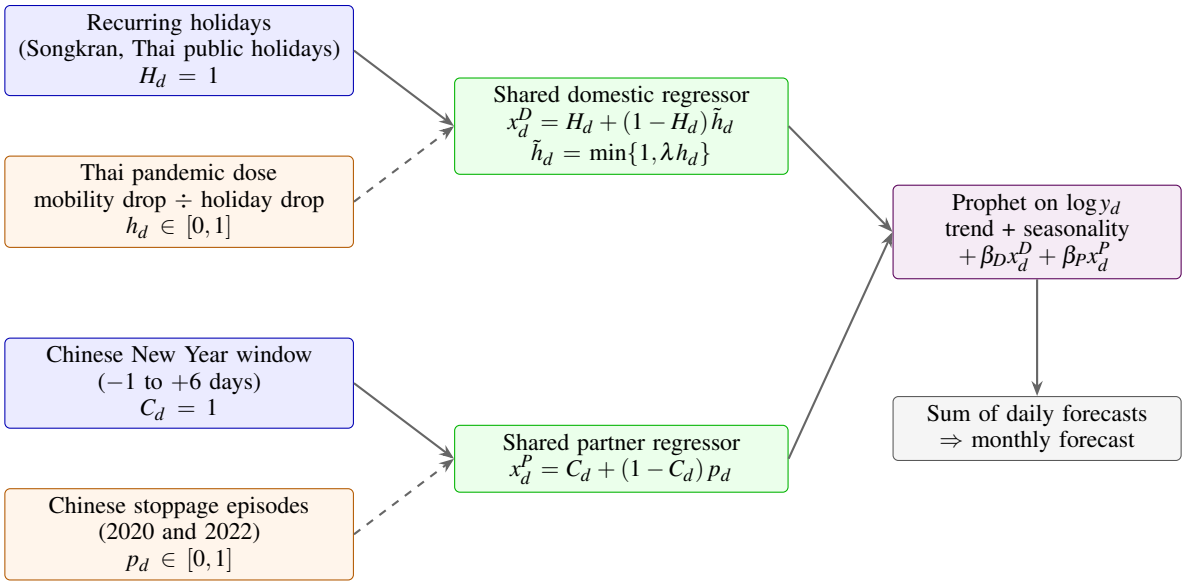


Figure 1. The HES framework. *Colour key:* blue boxes are the recurring holidays that identify the elasticities before the pandemic; orange boxes are the dose inputs measured during the pandemic; green boxes are the shared regressors that carry both; the violet box is the forecasting model and the grey box the monthly output. Solid arrows: the holiday indicators that identify β_D and β_P; dashed arrows: dose inputs that enter the same regressors during the pandemic.

3 The Holiday-Equivalent Stoppage framework

3.1 Setup and notation

Let d index calendar days and m months, let y_d denote the value of exports on day d, $Y_m = \sum_{d \in m} y_d$ the monthly total and D_m the number of days in month m. Let $H_d = 1$ if d is an *observed* Thai public holiday, including in-lieu and special holidays (for example, the 2020 Songkran holiday was postponed and replaced by 27 July, 4 September and 7 September 2020), and $C_d = 1$ if d lies in the Chinese-New-Year window $[\tau_y - 1, \tau_y + 6]$, where τ_y is the first day of the Spring Festival in year y. The daily model is

$$\log y_d \;=\; g(t_d) + s(t_d) + w(d) + \beta_D\, x_d^D + \beta_P\, x_d^P + D_d + u_d, \tag{1}$$

where g is a piecewise-linear trend, s a Fourier representation of yearly seasonality, w a day-of-week effect, D_d an unobserved non-stoppage shock (for instance, a collapse of foreign demand) and u_d a stationary disturbance.

3.2 Shared stoppage regressors

Definition 1 (Shared stoppage regressors)**.** Given a domestic dose $h_d \in [0,1]$, a partner dose $p_d \in [0,1]$ and a dose multiplier $\lambda > 0$,

$$x_d^D = H_d + (1 - H_d)\,\min\{1,\ \lambda h_d\}, \tag{2}$$

$$x_d^P = C_d + (1 - C_d)\,p_d. \tag{3}$$

The defining feature is that β_D and β_P multiply *both* the recurring holiday indicators and the pandemic doses. Before 2020 the doses are zero, so $x_d^D = H_d$ and $x_d^P = C_d$, and the elasticities are identified from about a decade of holidays. During the pandemic, the forecast requires only the doses; no new parameter is introduced. Figure 1 summarises the structure.

3.3 Measuring the dose: the Holiday-Equivalent Stoppage Index

Let $m_d = \max\{0, -\text{workplaces}_d/100\}$ be the proportional drop in visits to workplaces relative to the pre-pandemic baseline in the Google Community Mobility Reports, and let $\mathscr{H}^{wd}$ be the set of public holidays that fall on weekdays. The domestic dose is

$$h_d \;=\; \min\Big\{1,\ \frac{\tilde{m}_d}{\text{med}\{m_e : e \in \mathscr{H}^{wd}\}}\Big\}, \tag{4}$$

where $\tilde{m}_d$ is a seven-day centred moving average computed after removing holidays, so that holiday dips do not leak into neighbouring days. In the Thai data from 15 February 2020 to 19 August 2022 there are 57 weekday holidays with a median drop of 0.47 (Figure 3a); Songkran and New Year days are stronger, at 0.64–0.70. The average of h_d in April 2020 is 0.60: in terms of attendance at workplaces, a day of the April 2020 lockdown was equivalent to six tenths of a holiday.

Workplace attendance is not export production. Office work can move home while export-oriented factories, typically designated as essential, keep operating. We therefore model the *true* stoppage dose as

$$z_d = \kappa h_d (1 + \nu_d), \tag{5}$$

where $\kappa > 0$ is an unknown pass-through and ν_d a mean-zero, slowly varying proxy error. The multiplier λ in (2) is the forecaster's choice for κ.

For the partner channel, policy stringency is not a measure of stoppage. Under China's zero-COVID policy the OxCGRT index stayed between 0.6 and 0.8 for three years while most factories operated (Figure 3c). We therefore use an index of industrial-stoppage episodes (the national lockdown of late January to March 2020 and the Shanghai lockdown of April–May 2022) expressed in Chinese-New-Year units.

3.4 Identifying assumptions

Assumption A1 (Mechanism invariance)**.** The effect of stoppage on log exports is β_D per holiday-equivalent unit, irrespective of whether the stoppage is caused by a holiday or by a lockdown.

Assumption A2 (Dose measurability up to scale)**.** The true dose satisfies (5) with h_d observed, κ constant and $\mathbb{E}[\nu_d \mid h_d] = 0$.

Assumption A3 (Log-additivity)**.** Equation (1) holds, so stoppage enters additively on the log scale.

Assumption A4 (Non-stoppage shocks)**.** The non-stoppage shock D_d is unobserved. When $\mathbb{E}[D_d h_d] \neq 0$ over the forecast window, the optimal dose departs from κ (Proposition 1).

Assumption A1 is the substantive hypothesis of the paper. It is violated, for example, if long festivals have larger effects per day than single-day holidays; Section 6.5 studies this case and a two-tier specification that restores the assumption.

3.5 Estimation and forecasting

We estimate (1) with Prophet on $\log y_d$, using a piecewise-linear trend, yearly seasonality of Fourier order 10, weekly seasonality of order 3, and the shared regressors added with `standardize=False`. Because the regressors enter additively and unstandardised, the log forecast is exactly linear in the regressors,

$$\log \hat{y}_d = b_{0,d} + \hat{\beta}_D x_d^D + \hat{\beta}_P x_d^P, \tag{6}$$

where $b_{0,d}$ is the forecast with all regressors set to zero; we verified that the identity holds to within 10^{-15}. Dose scenarios can therefore be evaluated without re-estimating the model. Levels are recovered with the smearing estimator of Duan (1983), and monthly forecasts are sums of daily forecasts.

When only monthly data are available, the dependent variable is $\log(Y_m/D_m)$, the regressors are the monthly means $\bar{x}_m^D$ and $\bar{x}_m^P$ of the daily regressors together with the share of weekend days WK_m, and yearly seasonality has Fourier order 5.

4 Theoretical properties

This section collects the properties that govern when the transfer succeeds. Proofs are in Appendix A.

Remark 1 (One-off holidays do not transfer). In Prophet, each holiday name has its own coefficients with Gaussian priors centred at zero. If a "COVID" holiday is added that does not occur in the estimation window, its design columns are identically zero in that window, the likelihood does not depend on its coefficients, and the maximum a posteriori estimate equals the prior mode, zero. The forecast of an unobserved lockdown is therefore the no-lockdown forecast. Shared regressors avoid this by tying the lockdown coefficient to the holiday coefficient.

Proposition 1 (Transfer error and optimal dose)**.** *Under Assumptions A2–A4, on a non-holiday day with $\lambda h_d \leq 1$ and $z_d \leq 1$, the log forecast error of (6) is*

$$e_d = \beta_D(\kappa - \lambda)h_d + \beta_D \kappa h_d \nu_d + (\beta_D - \hat{\beta}_D)\lambda h_d + r_d, \tag{7}$$

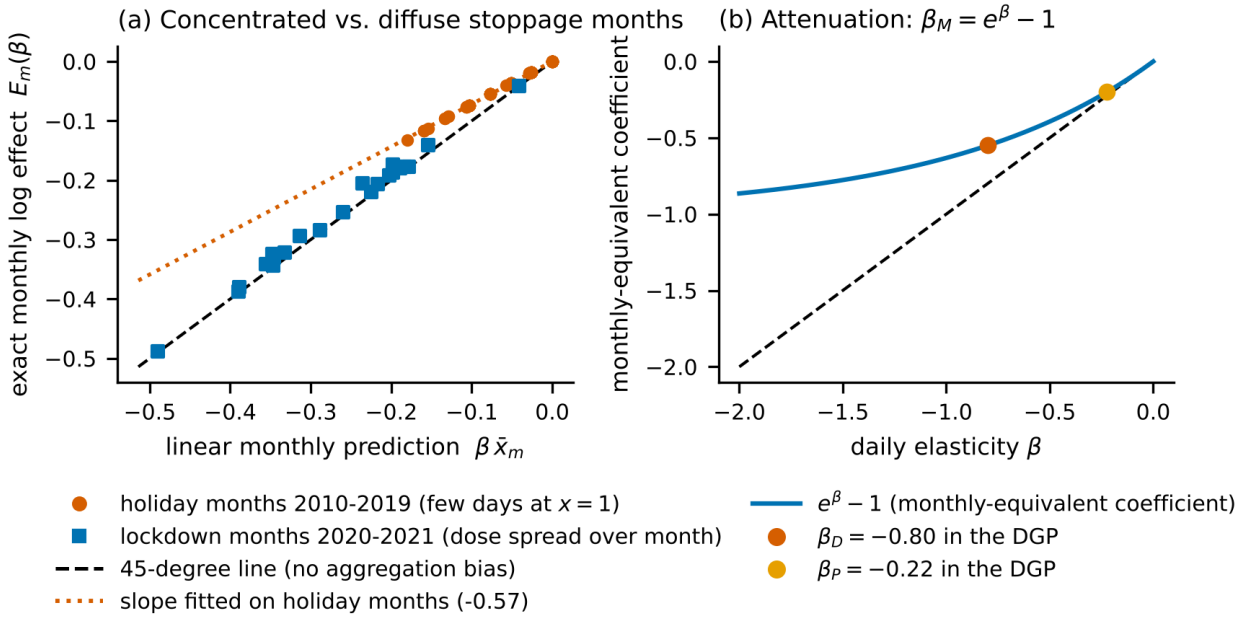


Figure 2. Aggregation bias (Proposition 2). (a) Exact monthly log effect against its linear approximation for the domestic channel: vermilion circles are holiday months 2010–2019, blue squares are lockdown months 2020–2021 with the S1 dose; the black dashed line is the 45-degree line and the vermilion dotted line the slope fitted on holiday months. Holiday months lie above the 45-degree line, lockdown months on it. (b) The monthly-equivalent coefficient $e^{\beta} - 1$ (blue) against the daily elasticity (black dashed: identity); the vermilion and orange points mark the domestic and partner elasticities of the data-generating process.

where r_d collects the errors of the other components and D_d. (a) If $\hat{\beta}_D = \beta_D$ and $\mathbb{E}[r_d] = 0$, then $\mathbb{E}[e_d] = \beta_D(\kappa - \lambda)h_d$, which vanishes if and only if $\lambda = \kappa$. (b) Over a window $\mathscr{W}$, the dose that minimises $\sum_{d\in\mathscr{W}} \mathbb{E}[e_d^2]$ is

$$\lambda^* = \kappa + \frac{\sum_{d\in\mathscr{W}} h_d \, \mathbb{E}[r_d]}{\beta_D \sum_{d\in\mathscr{W}} h_d^2}, \tag{8}$$

so a calibrated dose absorbs the projection of unmodelled shocks on the dose path. Because $\beta_D < 0$, a demand collapse that coincides with the lockdown ($\mathbb{E}[r_d] < 0$) pushes λ^* above κ. (c) The whole-month switch sets $\lambda h_d \equiv 1$ in selected months, yielding mean error $\beta_D(\kappa h_d - 1)$ there and $\beta_D \kappa h_d$ in the months not selected.

Proposition 2 (Aggregation bias)**.** *Let $w_d \geq 0$, $\sum_{d\in m} w_d = 1$, be the weights of day d in month m implied by the non-stoppage components, and let $E_m(\beta) = \log \sum_{d\in m} w_d e^{\beta x_d}$ be the exact effect of daily stoppage on the log of the monthly total. Then:*

- *(a) $E_m(\beta) \geq \beta \bar{x}_m$, with $\bar{x}_m = \sum_d w_d x_d$ and equality if and only if x_d is constant on the support of w;*
- *(b) $E_m(\beta) = \beta \bar{x}_m + \frac{1}{2}\beta^2 \operatorname{Var}_w(x) + O(\beta^3)$;*
- *(c) in a month in which a share q of days has $x_d = 1$ and the remainder $x_d = 0$, $E_m(\beta) = \log\{1 + q(e^{\beta} - 1)\}$; hence the least-squares slope of E_m on $\bar{x}_m$ across such months satisfies $\beta_M = e^{\beta} - 1 + O(\max q)$, and $|\beta_M| < |\beta|$ for $\beta < 0$;*
- *(d) in a month with a constant dose c, $E_m(\beta) = \beta c$, so the monthly forecast $\beta_M c$ understates the magnitude of the effect by $(1 + \beta - e^{\beta})|c|$.*

Holiday months are *concentrated* (a few days at one), whereas lockdown months are *diffuse* (a dose spread over the month). A monthly regression learns the attenuated coefficient from the former and applies it to the latter. For the Thai calendar and $\beta_D = -0.80$, the coefficient implied by exact aggregation is -0.57, close to the closed form $e^{\beta_D} - 1 = -0.55$. Figure 2 visualises the mechanism.

Corollary 1 (Aggregation correction)**.** *For holiday shares that are small, inverting Proposition 2(c) gives the closed-form correction $\hat{\beta}_D^{AC} = \log(1 + \hat{\beta}_M)$ for $\hat{\beta}_M > -1$, with delta-method variance $\operatorname{Var}(\hat{\beta}_D^{AC}) \approx \operatorname{Var}(\hat{\beta}_M)/(1 + \hat{\beta}_M)^2$. The*

exact minimum-distance version, $\hat{\beta}_D^{AC} = \arg\min_b \sum_m \{E_m(b) - \hat{\beta}_M \bar{x}_m\}^2$*, uses the actual holiday composition of each month and requires only the daily regressors, not daily exports.*

The correction removes the bias but inflates the variance by the factor $(1+\hat{\beta}_M)^{-2}$, which equals about 4.7 at $\hat{\beta}_M = -0.54$. Bias and variance are traded off, not eliminated.

Proposition 3 (Identification at monthly frequency)**.** *In the monthly regression of* $\log(Y_m/D_m)$ *on calendar-month effects, a trend and* $\bar{x}_m^D$*, the Frisch–Waugh–Lovell theorem gives* $\mathrm{Var}(\hat{\beta}_M) = \sigma_m^2/\sum_m \tilde{x}_m^2$*, where* $\tilde{x}_m$ *is the residual of* $\bar{x}_m^D$ *after projection on the calendar-month effects and the trend. Because Songkran falls on 13–15 April every year, its contribution to* $\bar{x}_m^D$ *is absorbed by the April effect, and* $\hat{\beta}_M$ *is identified only by year-to-year variation in the number of holidays per month (in-lieu days, lunar Buddhist holidays, special holidays and the 2020 postponement). In daily data,* β_D *is identified by contrasts between holidays and adjacent working days, which are nearly orthogonal to smooth seasonality.*

For 2010–2019 the standard deviation of $\bar{x}_m^D$ is 0.059, but after removing calendar-month means it falls to 0.029, roughly 0.9 of a holiday day per month. The corresponding daily residual standard deviation is 0.231.

Proposition 4 (Dose saturation in tree ensembles)**.** *Let* f *be an ensemble of regression trees trained on data in which a feature* x *takes only the values 0 and 1. Every split on* x *uses a threshold* $c \in (0,1)$*, and with histogram binning all trees share the same threshold. Holding the other features fixed,* f *takes only the two values* $f(0)$ *and* $f(1)$*, so* $f(\lambda h) = f(1)$ *for every dose* $\lambda h > c$*. If* x *is continuous with training support* $[0,a]$*, then* $f(x) = f(a)$ *for* $x > a$*.*

In LightGBM the threshold for a binary feature is placed at 10^{-35}, which we verified in the fitted trees: any positive dose is treated as a full holiday.

5 Simulation design

We follow the ADEMP structure of Morris et al. (2019): aims, data-generating mechanisms, estimands, methods and performance measures.

5.1 Aims

The study evaluates whether stoppage elasticities identified from recurring holidays transfer to lockdowns (RQ1), how accuracy depends on the dose scale and its calibration (RQ2), what monthly data lose (RQ3), whether tree ensembles can transfer (RQ4) and how the shock should be handled afterwards (RQ5). Simulation is necessary because the true elasticities, pass-through and doses are unknown in official data.

5.2 Data-generating mechanisms

Real inputs. The holiday calendar is the observed Thai calendar from the `holidays` package (python-holidays contributors, 2025), excluding National Children's Day, which is not a day off. Chinese-New-Year dates are taken from the Chinese calendar in the same package. Daily national stringency indices for Thailand and China come from OxCGRT (Hale et al., 2021), and daily workplace mobility for Thailand from the Google Community Mobility Reports (Google LLC, 2022). Figure 3 displays these inputs.

Data-generating process. Daily exports from 1 January 2010 to 31 December 2024 follow (1) with the parameters in Table 1. The level and trend place annual exports in the range of USD 200–300 billion, the order of magnitude of Thai merchandise exports. Holidays reduce exports by 55% ($\beta_D = \log 0.45$) and each day of the Chinese-New-Year window by 20% ($\beta_P = \log 0.80$). The true pandemic doses follow (5) with an AR(1) proxy error ν_d (persistence 0.97, stationary standard deviation 0.15), so mobility is an imperfect proxy even in the best case.

Scenarios. In *S1 (ideal transfer)* the pass-through is $\kappa = 1$ and there is no demand shock: Assumptions A1–A3 hold and HESI is correctly scaled up to ν_d. In *S2 (realistic)* the pass-through is $\kappa = 0.5$, so mobility overstates the stoppage two-fold, and a global demand shock that HESI cannot see lowers exports by 8% in April–June 2020 and raises them by 7% in 2021 before fading by mid-2022. Each replication uses the same random numbers in both scenarios (common random numbers), so differences between scenarios are not driven by noise. Figure 4 shows one replication.

Robustness mechanisms. *R1 (heterogeneous holidays)* replaces the common holiday effect with $\log 0.35 = -1.05$ for festival days (Songkran and New Year) and $\log 0.55 = -0.60$ for ordinary holidays, with the pandemic dose

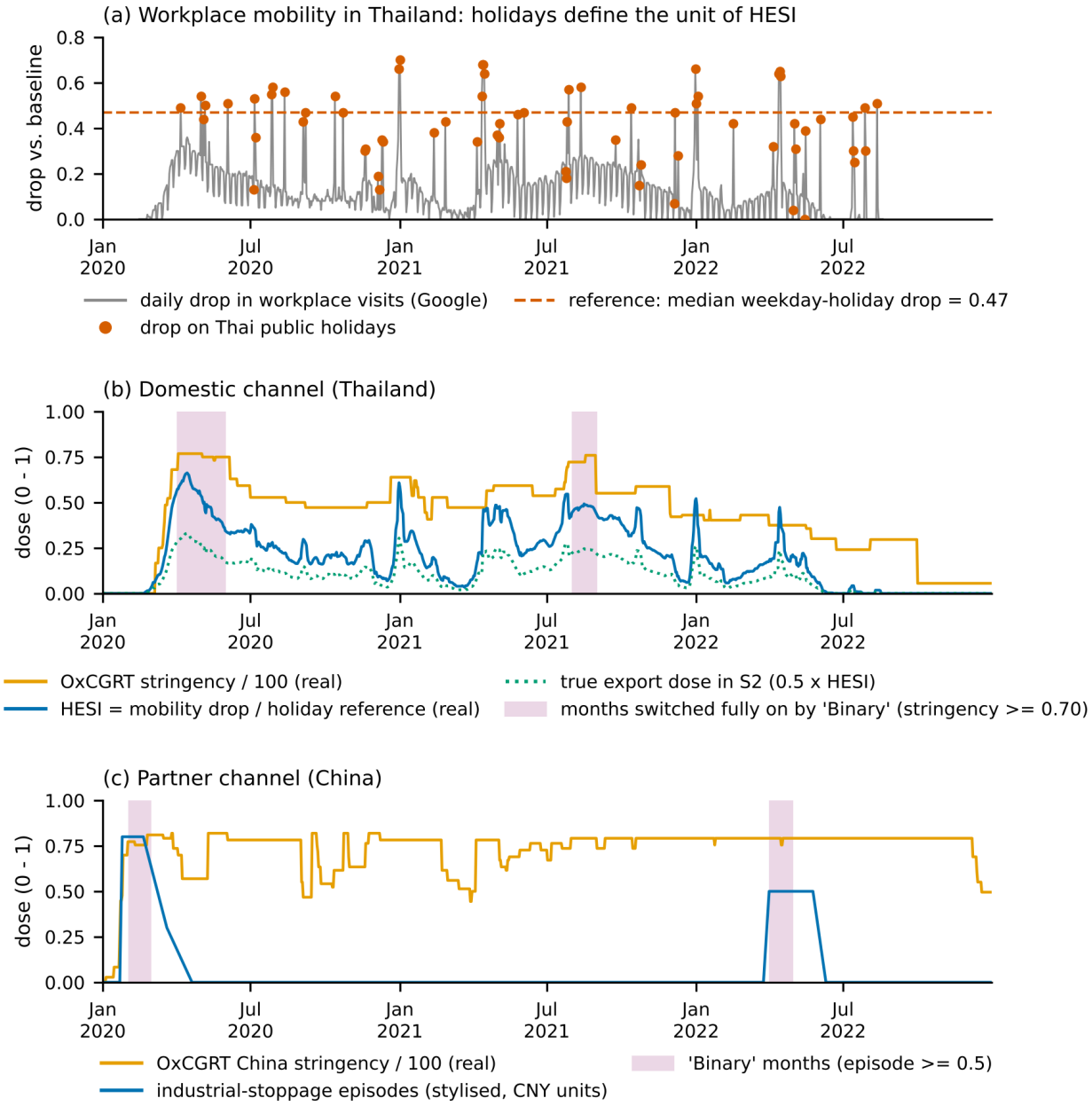


Figure 3. Real pandemic inputs. (a) Daily drop in workplace visits in Thailand (grey line) and the drop on public holidays (vermilion points); the vermilion dashed line is the median drop on weekday holidays, which defines one holiday unit of HESI. (b) Domestic channel: OxCGRT stringency (orange), HESI (blue) and the true export dose of scenario S2 (green dotted); lilac bands mark the months that the Binary variant switches fully on. (c) Partner channel: OxCGRT stringency for China (orange) stays high for three years, whereas industrial-stoppage episodes (blue) are short; lilac bands mark the Binary months.

Table 1. Parameters of the daily data-generating process (log scale).

Component	Specification
Level	USD 560 million per day in 2010 (about USD 200 billion per year)
Trend	$+10\%$/yr (2010–11), -1%/yr (2012–15), $+2.5\%$/yr (2016–19), $+3\%$/yr (2020–22), $+2\%$/yr (2023–24)
Yearly seasonality	three Fourier harmonics, amplitude about $\pm 6\%$
Day of week	Monday–Friday $+0.02$ to $+0.04$; Saturday -0.30; Sunday -0.55
Holiday elasticity β_D	$\log 0.45 = -0.80$ per holiday-equivalent day
Partner elasticity β_P	$\log 0.80 = -0.22$ per day in the Chinese-New-Year window
Pandemic doses	domestic $\kappa h_d(1+\nu_d)$; partner $p_d(1+\nu'_d)$; ν, ν' AR(1), $\rho = 0.97$, s.d. 0.15
Disturbances	daily AR(1) $\rho = 0.985$ (stationary s.d. 0.046); monthly AR(1) $\rho = 0.85$ (innovation s.d. 0.015); i.i.d. s.d. 0.12
S1	$\kappa = 1$, no demand shock
S2	$\kappa = 0.5$; demand -8% in Apr–Jun 2020, $+7\%$ in 2021, zero after mid-2022

Table 2. Methods and dose variants.

Method	Specification
Seasonal naive	value of the same month in the last training year
Prophet daily (vanilla)	Prophet on $\log y_d$ without calendar regressors, summed to months
Prophet daily (P-D)	Prophet on $\log y_d$ with shared regressors x_d^D, x_d^P
Prophet monthly (P-M)	Prophet on $\log(Y_m/D_m)$ with $\bar{x}_m^D$, $\bar{x}_m^P$, WK_m (Fourier order 5, 15 changepoints)
P-M + AC	P-M with the aggregation correction of Corollary 1
SARIMAX	airline model $(0,1,1)(0,1,1)_{12}$ on $\log(Y_m/D_m)$ with the P-M regressors (Box et al., 2015)
LightGBM daily / monthly	OLS log-linear trend plus LightGBM on the detrended residual; daily features: day of week, day of year, x_d^D, x_d^P; monthly features: month, $\bar{x}_m^D$, $\bar{x}_m^P$, WK_m
Dose variants in the shock period (E1)	
Off	holidays only, no pandemic dose
Binary whole-month switch	$x^D = 1$ on every day of months with mean stringency ≥ 0.70; $x^P = 1$ in months with episode index ≥ 0.5
Raw stringency	OxCGRT stringency/100 for Thailand and China used as doses
HESI ($\lambda = 1$)	mobility-based HESI for Thailand; episode index for China
HESI-cal ($\hat{\lambda}$)	$\hat{\lambda} \in \{0, 0.05, \ldots, 3\}$ minimising squared log errors in March–April 2020; evaluated May 2020–December 2021
Oracle (κ known)	$\lambda = \kappa$; the proxy error ν_d and the demand shock remain unknown
Treatments of the pandemic in the training sample (E2)	
COVID ignored	holiday regressors only
Mask COVID window	observations from 20 January 2020 to 30 September 2022 removed
Separate COVID regressors	HESI and episode index enter with their own coefficients
Binary, raw stringency, HESI	the corresponding shared regressors enter the training sample

measured in ordinary-holiday units. *R2 (real-time dose)* replaces the centred smoothing in (4) with a trailing seven-day window available in real time.

5.3 Estimands

The primary estimand is the monthly export total in each test window, evaluated by MAPE and by the mean absolute scaled error (MASE) of Hyndman and Koehler (2006), scaled by the in-sample seasonal-naive error. Secondary estimands are the elasticities β_D and β_P (bias, standard deviation and root mean squared error across replications) and the calibrated dose $\hat{\lambda}$ relative to κ.

5.4 Methods

Table 2 lists the forecasting methods and the dose variants. All hyperparameters are fixed in advance; nothing is tuned on test data.

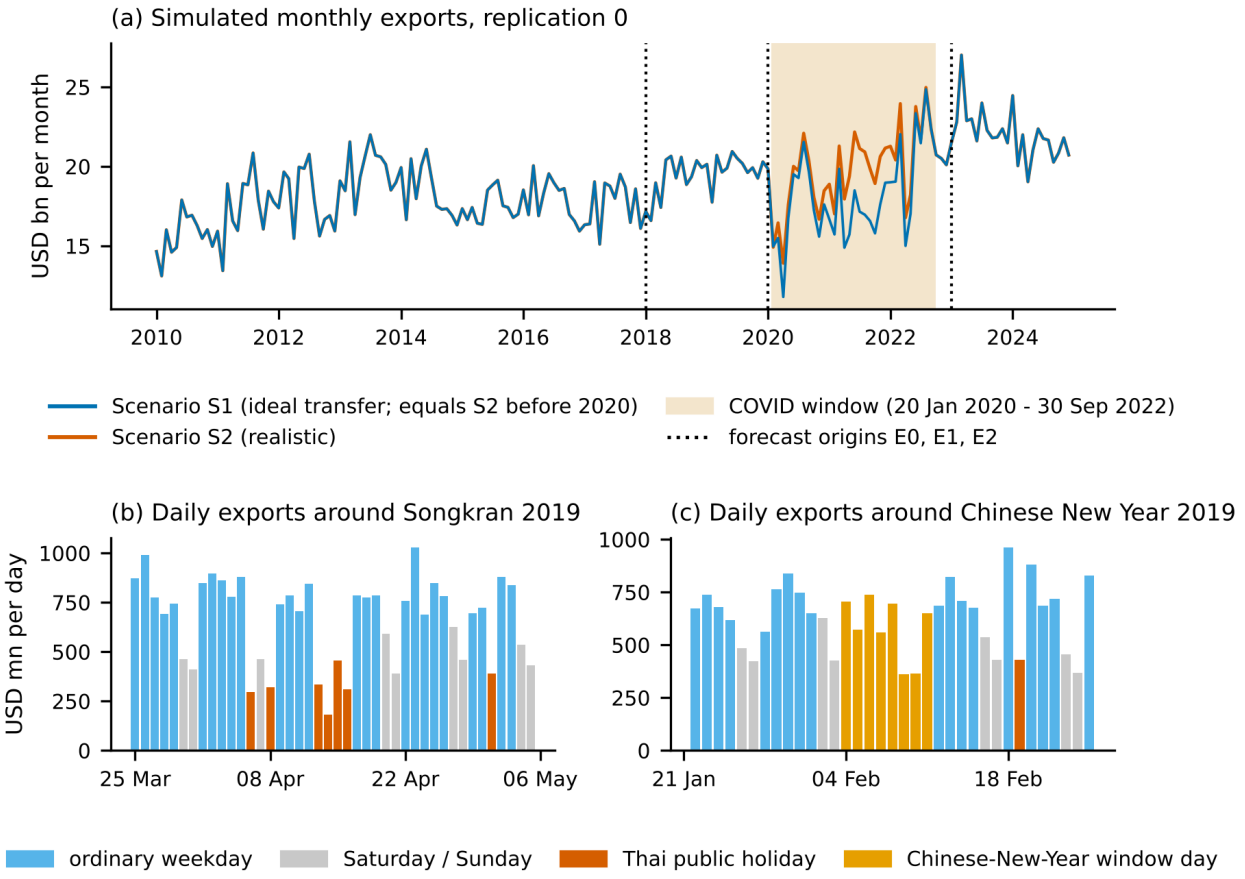


Figure 4. Simulated data, replication 0. (a) Monthly totals in S1 (blue) and S2 (vermilion), which coincide before 2020 because they share the same random numbers; the tan band is the pandemic window and the dotted black lines are the forecast origins of experiments E0, E1 and E2. (b)–(c) Daily values around Songkran and the Chinese New Year 2019, coloured by day type: sky blue for ordinary weekdays, light grey for weekends, vermilion for Thai public holidays and orange for days in the Chinese-New-Year window.

5.5 Experiments

E0 (normal period): training to December 2017, forecasts for January 2018–December 2019. **E1 (shock period)**: training to December 2019, forecasts for January 2020–December 2021. The doses observed during the forecast window are supplied to the models, so E1 evaluates *scenario-conditional* forecasts (equivalently, nowcasts conditional on observed mobility), not ex-ante forecasts. Accuracy is reported over the full 24 months and over May 2020–December 2021, the window that follows the calibration months of HESI-cal. **E2 (aftermath)**: training to December 2022, including the pandemic, and forecasts for January 2023–December 2024 with no pandemic doses.

5.6 Performance measures and inference

For each method we report the mean across 30 replications and its Monte Carlo standard error. Because all methods are applied to the same replications, comparisons are paired: for each comparison we report the mean paired difference in MAPE with a 95% t-interval, the share of replications in which the first method is more accurate, and the Wilcoxon signed-rank test (Wilcoxon, 1945), with Holm's correction across the family of comparisons. In this setting the replication, not the month, is the unit of inference, so we do not apply Diebold–Mariano tests (Diebold and Mariano, 1995; Harvey et al., 1997) within replications; they are part of the protocol for official data (Section 7.2).

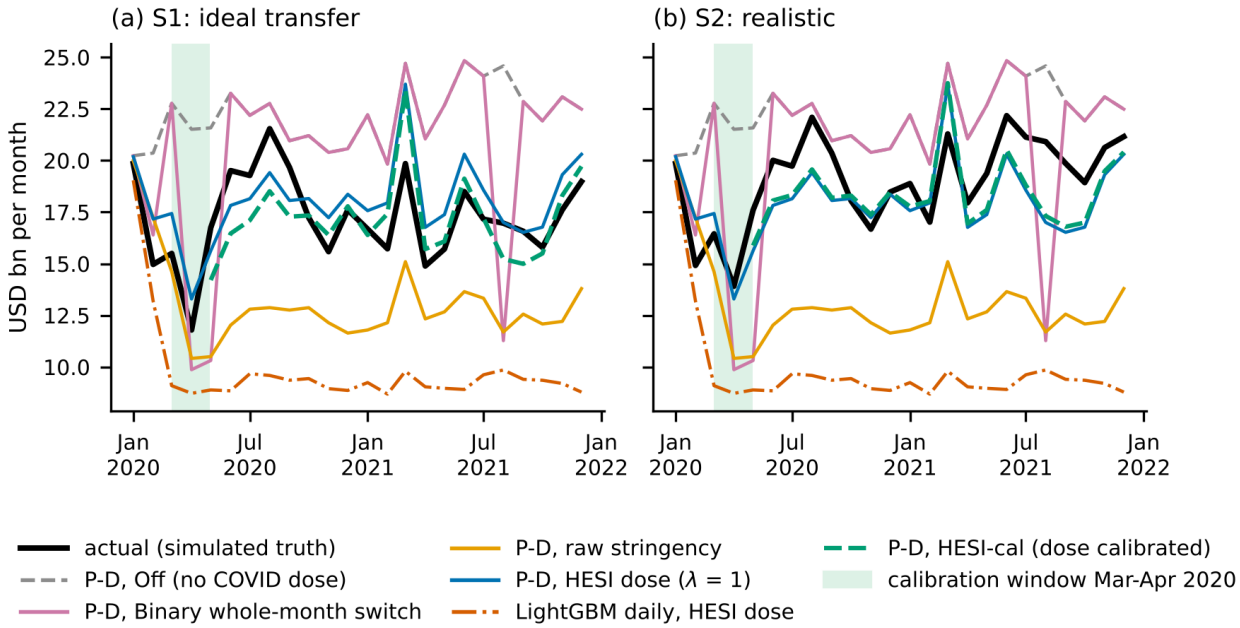


Figure 5. Forecasts for the shock period from December 2019 (replication 0). Black: actual simulated totals. Grey dashed: P-D without a dose. Lilac: P-D with the Binary whole-month switch. Orange: P-D with raw stringency. Blue: P-D with HESI ($\lambda = 1$). Green dashed: P-D with the calibrated dose, shown from May 2020. Vermilion dash-dot: LightGBM daily with HESI. The mint band marks the calibration window, March–April 2020.

6 Results

6.1 Transfer during the shock (E1)

Table 3 reports the accuracy of all methods in the shock period, and Figure 5 shows the forecast paths of one replication. In S1, daily Prophet with HESI reduces the MAPE from 26.0% without a dose to 7.8%, and it is more accurate than the no-dose forecast in 100% of replications. The model never observes a pandemic day: the entire improvement comes from the elasticity identified on holidays before 2020. This answers RQ1 affirmatively under the assumptions of S1.

The whole-month switch reduces the MAPE only to 22.7%. Consistent with Proposition 1(c), it overshoots in the three months that it switches on, where it predicts a 55% loss for the whole month, and it leaves the remaining months of partial stoppage unadjusted. Raw stringency raises the MAPE to 27.2%, mainly because China's stringency index assigns a dose of 0.6–0.8 to a partner that was largely operating.

In S2, where mobility overstates the stoppage two-fold and a demand shock is present, HESI with $\lambda = 1$ improves on the no-dose forecast on average (10.9% against 14.7%) but not consistently: it wins in 70% of replications, and the difference is not significant after Holm's correction (adjusted $p = 0.155$). It also falls short of the oracle (8.3%). Calibrating the dose on March–April 2020 lowers the MAPE over May 2020–December 2021 to 8.8%, against 13.0% without a dose and 8.6% for the oracle. The calibrated dose is significantly better than the uncalibrated one (adjusted $p = <0.001$), whereas its advantage over the no-dose forecast is significant only before adjustment (adjusted $p = 0.093$). Figure 6 shows the corresponding distributions across replications.

One result requires care. In S2 the monthly model with $\lambda = 1$ is more accurate than the daily model (8.6% against 10.9%). This is not an advantage of monthly data but a cancellation of two errors: the attenuated monthly coefficient (Proposition 2) offsets the two-fold overstatement of the dose. When the dose is correctly scaled (S1), the ranking reverses (7.8% against 11.0%).

6.2 Dose sensitivity and real-time calibration

Figure 7 traces the MAPE as a function of the dose multiplier λ. For daily Prophet the curve is U-shaped with its minimum at $\lambda = 1.05$ in S1 (true $\kappa = 1$) and at $\lambda = 0.60$ in S2 (true $\kappa = 0.5$), as Proposition 1 predicts. For the monthly models the minimum shifts to the right, to 1.55 for Prophet and 1.45 for SARIMAX in S1: an attenuated

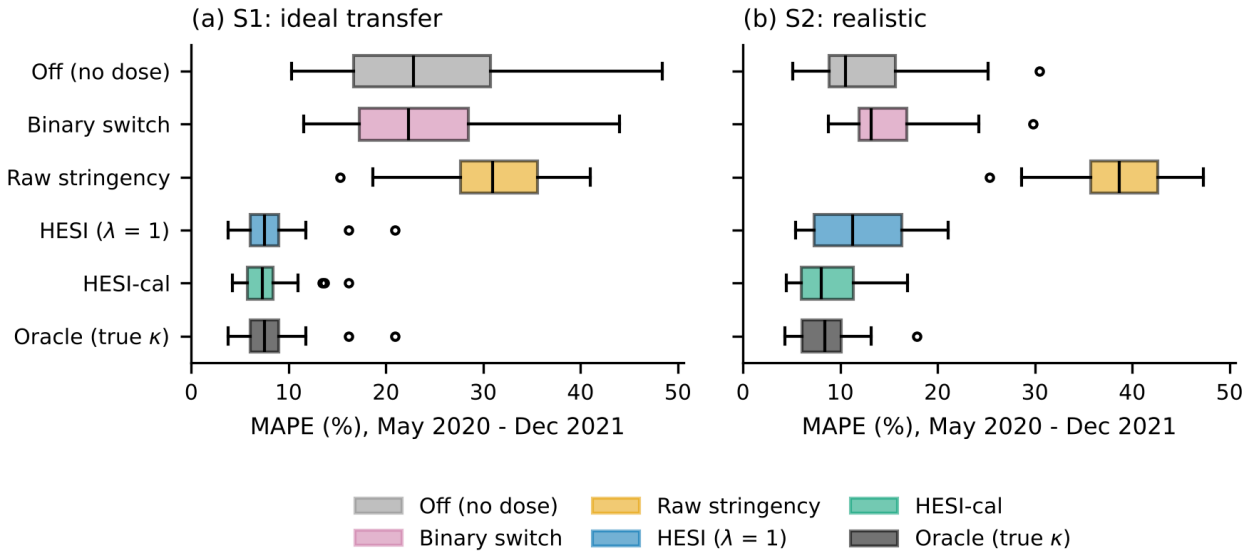


Figure 6. Distribution of the MAPE of daily Prophet across 30 replications, May 2020–December 2021, by dose variant. Box colours: grey for no dose, lilac for the Binary switch, orange for raw stringency, blue for HESI with $\lambda = 1$, green for the calibrated dose and dark grey for the oracle. Boxes show the interquartile range, black lines the median, whiskers 1.5 times the interquartile range and circles the remaining replications.

elasticity requires an inflated dose. The aggregation correction moves the minimum back to 1.05 in S1 and 0.55 in S2, but its larger variance keeps its minimum error above that of the daily model.

Table 4 reports the doses selected by real-time calibration. Daily Prophet selects $\hat{\lambda} = 1.07$ in S1 and 0.72 in S2. The upward deviation in S2 is the absorption effect of Proposition 1(b): evaluating (8) with the true demand path over the non-holiday days of March–April 2020 gives $\lambda^* = 0.66$, close to the mean calibrated value. Calibration on exports alone cannot distinguish stoppage from a coincident demand collapse.

6.3 What monthly data lose: aggregation and identification

Figure 8 and Table 5 compare the elasticities estimated from the pre-pandemic sample with their true values and with the predictions of Section 4. Daily Prophet recovers β_D almost exactly (-0.80, standard deviation 0.01). Monthly Prophet and SARIMAX both yield -0.54, close to the aggregation-implied -0.57 and the closed form -0.55: the attenuation is explained by Proposition 2. Their standard deviations, 0.13 and 0.14, are more than ten times that of the daily estimate, as Proposition 3 implies. The aggregation correction raises the mean to -0.76 and the standard deviation to 0.27; the closed form of Corollary 1 predicts -0.78 and 0.28. For the partner channel, the elasticity is small and the Chinese New Year moves between January and February, so monthly estimation is nearly unbiased (-0.20 against -0.22; closed form -0.20) but noisy (standard deviation 0.09 against 0.02). These results answer RQ3: monthly data lose precision through identification and accuracy through aggregation; the correction recovers the latter but not the former.

6.4 Tree ensembles saturate

Daily LightGBM observes the stoppage regressor only at 0 and 1 before 2020 and splits it at 10^{-35} in every tree. Any positive dose is therefore treated as a full holiday, and the MAPE rises to 39.7% in S1 and 45.6% in S2; real-time calibration correspondingly selects a dose at or near zero (mean 0.01). Monthly LightGBM observes monthly holiday shares of at most about 0.2, whereas lockdown months reach 0.6; it predicts the effect of the most holiday-intensive month it has seen and improves the MAPE only from 21.6% to 16.6% in S1. These outcomes follow from Proposition 4 and do not depend on tuning, which answers RQ4.

Table 3. Shock period (E1): MAPE (%), mean over 30 replications with Monte Carlo standard errors in parentheses. Bold: lowest value in each column excluding the oracle. "Jan–Dec" denotes January 2020–December 2021; "May–Dec" denotes May 2020–December 2021.

		S1: ideal transfer		S2: realistic	
Method	Dose variant	Jan–Dec	May–Dec	Jan–Dec	May–Dec
Seasonal naive	–	20.5 (0.7)	19.8 (0.8)	11.1 (0.5)	9.9 (0.5)
Prophet daily	No calendar regressors	26.4 (1.3)	26.4 (1.5)	14.7 (0.9)	13.8 (1.0)
Prophet daily (P-D)	Off (no dose)	26.0 (1.5)	24.9 (1.7)	14.7 (1.0)	13.0 (1.1)
	Binary whole-month switch	22.7 (1.2)	23.9 (1.4)	15.5 (0.7)	15.0 (0.9)
	Raw stringency	27.2 (0.9)	30.5 (1.1)	34.4 (0.8)	38.3 (0.9)
	HESI ($\lambda = 1$)	**7.8 (0.5)**	8.1 (0.6)	10.9 (0.7)	11.6 (0.8)
	HESI-cal ($\hat{\lambda}$)	–	**7.7 (0.5)**	–	8.8 (0.6)
	Oracle (κ known)	7.8 (0.5)	8.1 (0.6)	8.3 (0.5)	8.6 (0.6)
Prophet monthly (P-M)	Off (no dose)	26.3 (1.3)	25.2 (1.5)	14.5 (0.9)	12.7 (1.0)
	Binary whole-month switch	21.2 (1.1)	22.6 (1.2)	13.2 (0.6)	13.1 (0.8)
	Raw stringency	17.2 (1.1)	19.2 (1.3)	24.3 (1.2)	27.2 (1.4)
	HESI ($\lambda = 1$)	11.0 (1.1)	11.1 (1.2)	8.6 (0.5)	9.0 (0.6)
	HESI-cal ($\hat{\lambda}$)	–	8.9 (0.7)	–	9.0 (0.6)
	Oracle (κ known)	11.0 (1.1)	11.1 (1.2)	9.3 (0.6)	9.0 (0.7)
P-M + AC	Binary whole-month switch	22.8 (1.0)	23.9 (1.2)	15.0 (0.7)	14.4 (0.8)
	Raw stringency	25.8 (1.8)	28.8 (2.0)	32.8 (1.7)	36.5 (1.8)
	HESI ($\lambda = 1$)	9.8 (1.0)	10.1 (1.1)	11.3 (0.9)	12.0 (1.0)
	HESI-cal ($\hat{\lambda}$)	–	8.0 (0.5)	–	8.6 (0.5)
	Oracle (κ known)	9.8 (1.0)	10.1 (1.1)	8.9 (0.6)	9.0 (0.6)
SARIMAX	Off (no dose)	24.8 (1.2)	23.7 (1.4)	13.4 (0.8)	11.5 (1.0)
	Binary whole-month switch	19.9 (1.0)	21.3 (1.2)	12.3 (0.6)	12.1 (0.7)
	Raw stringency	18.1 (1.2)	20.4 (1.4)	25.4 (1.3)	28.6 (1.4)
	HESI ($\lambda = 1$)	9.9 (1.0)	10.0 (1.1)	**8.5 (0.5)**	9.1 (0.5)
	HESI-cal ($\hat{\lambda}$)	–	8.4 (0.7)	–	8.6 (0.5)
	Oracle (κ known)	9.9 (1.0)	10.0 (1.1)	8.5 (0.6)	8.3 (0.7)
LightGBM daily	Off (no dose)	21.7 (0.8)	20.2 (0.9)	11.2 (0.5)	9.1 (0.5)
	Binary whole-month switch	19.4 (0.5)	19.9 (0.7)	12.9 (0.3)	11.9 (0.4)
	Raw stringency	40.4 (0.4)	44.4 (0.5)	46.2 (0.4)	50.7 (0.4)
	HESI ($\lambda = 1$)	39.7 (0.4)	43.1 (0.5)	45.6 (0.4)	49.5 (0.4)
	HESI-cal ($\hat{\lambda}$)	–	22.3 (1.5)	–	9.1 (0.5)
	Oracle (κ known)	39.7 (0.4)	43.1 (0.5)	45.6 (0.4)	49.5 (0.4)
LightGBM monthly	Off (no dose)	21.6 (0.8)	19.7 (0.9)	11.5 (0.5)	9.1 (0.5)
	Binary whole-month switch	20.0 (0.7)	18.5 (0.8)	10.3 (0.4)	8.3 (0.4)
	Raw stringency	15.5 (0.7)	14.3 (0.7)	8.9 (0.3)	**7.8 (0.3)**
	HESI ($\lambda = 1$)	16.6 (0.7)	15.5 (0.8)	9.2 (0.3)	8.0 (0.3)
	HESI-cal ($\hat{\lambda}$)	–	15.8 (0.8)	–	8.4 (0.3)
	Oracle (κ known)	16.6 (0.7)	15.5 (0.8)	9.3 (0.3)	8.1 (0.3)

6.5 Robustness: heterogeneous holidays and real-time doses

Table 6 reports the robustness experiments. Under R1, the one-tier specification pools festival and ordinary holidays and estimates -0.71, between the ordinary-holiday value (-0.60) and the festival value (-1.05). Because the dose is measured in ordinary-holiday units, the pooled elasticity acts as a dose multiplier of about 1.19. The two-tier specification, with a separate festival indicator and the dose carried by the ordinary-holiday regressor, recovers both elasticities (-0.60 and -1.04). The accuracy gain is negligible in S1 (7.5% against 7.5%), where the error curve is flat near its minimum, and larger in S2 (9.5% against 11.0%), where the pooled elasticity compounds the overstated dose. Under R2, replacing the centred smoothing with a trailing window available in real time changes the MAPE of daily Prophet from 7.8% to 7.9% in S1 and from 10.9% to 11.0% in S2, so the results in Table 3 do not rely on future mobility data within the smoothing window.

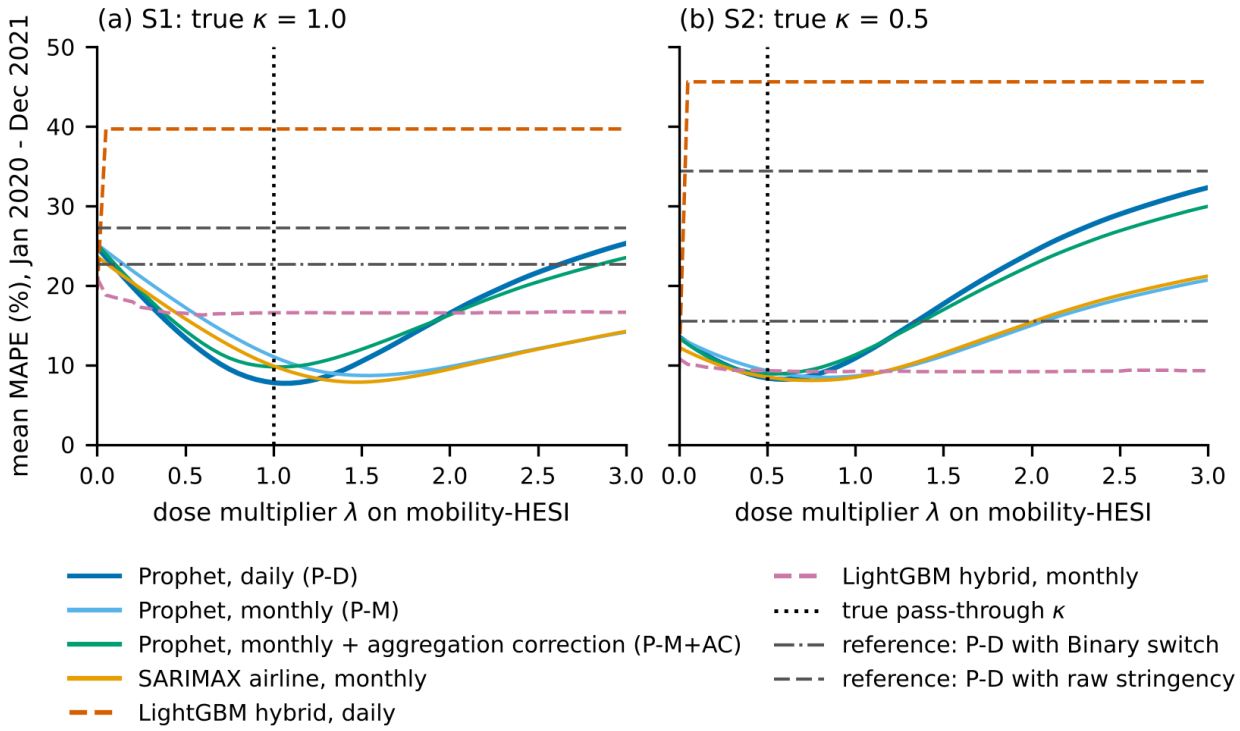


Figure 7. Mean MAPE over January 2020–December 2021 as a function of the dose multiplier λ applied to HESI. Line colours identify methods: blue for daily Prophet, sky blue for monthly Prophet, green for monthly Prophet with aggregation correction, orange for SARIMAX, vermilion dashed for daily LightGBM and lilac dashed for monthly LightGBM. The black dotted vertical line is the true pass-through κ; the grey dash-dot and dashed horizontal lines are daily Prophet with the Binary switch and with raw stringency.

Table 4. Calibrated dose $\hat{\lambda}$ selected on March–April 2020 (mean and standard deviation across replications). The true pass-through is $\kappa = 1$ in S1 and $\kappa = 0.5$ in S2; the last row reports the optimal dose implied by Proposition 1(b) for the calibration window.

	S1		S2	
Method	Mean	S.D.	Mean	S.D.
Prophet daily (P-D)	1.07	0.20	0.72	0.18
Prophet monthly (P-M)	1.87	0.67	1.19	0.53
P-M + AC	1.24	0.51	0.83	0.35
SARIMAX	1.77	0.56	1.12	0.44
LightGBM daily	0.01	0.02	0.00	0.00
LightGBM monthly	0.44	0.26	0.45	0.26
Prop. 1(b) optimum for P-D	1.00	–	0.66	–

6.6 Forecasting after the shock (E2)

Table 7 and Figure 9 report forecasts for 2023–2024 from models trained through the pandemic. Ignoring the pandemic gives a daily-Prophet MAPE of 8.8% in S1 and 10.8% in S2. Including HESI in the training sample lowers it to 6.2% and 8.4%, and masking the pandemic window lowers it to 6.1% and 6.1%. Masking is as accurate as HESI in S1 and clearly more accurate in S2, where the overstated dose and the unmodelled rebound leak into the trend. SARIMAX, which adapts its level to the most recent observations, has the lowest mean MAPE in both scenarios (5.7% and 5.7%), although its advantage over masked daily Prophet is not statistically significant (adjusted $p = 0.474$ and 0.474). Raw stringency is harmful (30.4% and 34.8%): the Chinese index stays high throughout the pandemic part of the training sample and drops to zero in the forecast period, so the model forecasts an immediate jump in exports. For RQ5, the evidence favours simple treatments—masking, or models that adapt their level—over structural treatments of the shock once it has ended.

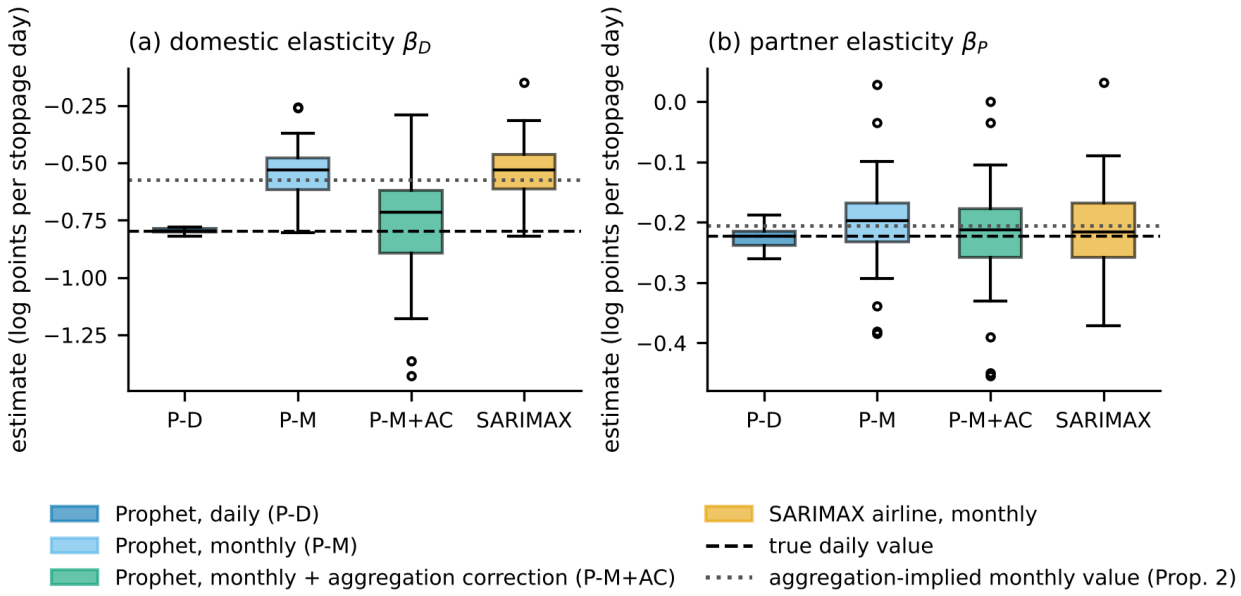


Figure 8. Elasticities estimated from data up to December 2019 across 30 replications. Box colours identify estimators: blue for daily Prophet, sky blue for monthly Prophet, green for monthly Prophet with aggregation correction and orange for SARIMAX. The black dashed line is the true daily value; the grey dotted line is the monthly value implied by exact aggregation (Proposition 2).

Table 5. Estimated elasticities (E1 fits, data to December 2019): mean, standard deviation and root mean squared error with respect to the true daily value, together with theoretical benchmarks.

	Domestic β_D			Partner β_P		
Estimator	Mean	S.D.	RMSE	Mean	S.D.	RMSE
Prophet daily (P-D)	−0.796	0.010	0.010	−0.225	0.017	0.017
Prophet monthly (P-M)	−0.540	0.131	0.289	−0.203	0.086	0.086
P-M + AC	−0.764	0.268	0.266	−0.225	0.099	0.097
SARIMAX	−0.538	0.137	0.294	−0.215	0.087	0.086
True daily value	−0.799			−0.223		
Exact aggregation (Prop. 2)	−0.573			−0.206		
Closed form $e^{\beta} - 1$	−0.550			−0.200		
AC closed form, Cor. 1 (delta s.d.)	−0.777	0.285		−0.227	0.107	

6.7 Normal times (E0)

In the absence of a shock (Table 8), daily LightGBM is the most accurate method (4.7%), followed by monthly LightGBM (5.5%) and SARIMAX (5.9%). Daily Prophet with holiday regressors (7.3%) is close to the seasonal naive benchmark (7.5%), and the holiday regressors improve it only marginally (from 7.4%). The value of HES lies in transfer during shocks, not in the general accuracy of Prophet.

6.8 Paired inference

Table 9 collects the paired comparisons behind the conclusions above. In S1, the contrasts that establish transfer—HESI against no dose, against the whole-month switch and against raw stringency, and the calibrated dose against no dose—are large relative to their confidence intervals and remain significant after Holm's correction. So do the advantage of daily Prophet over daily LightGBM in both scenarios, the advantage of masking or HESI over ignoring the pandemic in E2, and the advantage of daily LightGBM in normal times. In S2 the evidence for transfer is weaker: HESI and the calibrated dose improve on the no-dose forecast on average, but neither improvement survives Holm's correction (adjusted p = 0.155 and 0.093), whereas calibration significantly improves on the uncalibrated dose (adjusted p = <0.001). Several secondary contrasts are not significant after adjustment, including daily Prophet against

Table 6. Robustness experiments (E1, daily Prophet, 30 replications): MAPE (%) over January 2020–December 2021 with Monte Carlo standard errors, and estimated elasticities.

Specification	Dose variant	S1	S2
Panel A. R1, heterogeneous holidays: MAPE (%)			
One-tier (pooled holidays)	Off (no dose)	18.7 (1.3)	11.6 (0.8)
	HESI ($\lambda = 1$)	7.5 (0.5)	11.0 (0.7)
	Oracle (κ known)	7.5 (0.5)	8.1 (0.5)
Two-tier (festival indicator)	Off (no dose)	18.6 (1.3)	11.6 (0.8)
	HESI ($\lambda = 1$)	7.5 (0.5)	9.5 (0.6)
	Oracle (κ known)	7.5 (0.5)	8.3 (0.5)
Panel B. R1 elasticities: mean (s.d.) across replications			
One-tier pooled $\hat{\beta}$		-0.712 (0.010)	
Two-tier ordinary $\hat{\beta}_O$ (true -0.598)		-0.596 (0.012)	
Two-tier festival $\hat{\beta}_F$ (true -1.050)		-1.043 (0.018)	
Panel C. R2, real-time dose: MAPE (%)			
Centred smoothing	HESI ($\lambda = 1$)	7.8 (0.5)	10.9 (0.7)
Trailing smoothing	HESI ($\lambda = 1$)	7.9 (0.5)	11.0 (0.7)
Centred smoothing	Oracle (κ known)	7.8 (0.5)	8.3 (0.5)
Trailing smoothing	Oracle (κ known)	7.9 (0.5)	8.4 (0.5)

Table 7. Aftermath (E2, January 2023–December 2024): MAPE (%) with Monte Carlo standard errors and mean MASE. Bold: lowest MAPE in each scenario.

		S1		S2	
Method	Treatment of the pandemic	MAPE	MASE	MAPE	MASE
Seasonal naive	–	12.4 (0.4)	1.57	9.5 (0.4)	1.33
Prophet daily	No calendar regressors	8.0 (0.3)	1.00	8.9 (0.6)	1.21
	COVID ignored	8.8 (0.6)	1.09	10.8 (0.8)	1.46
	Mask COVID window	6.1 (0.3)	0.76	6.1 (0.3)	0.84
	Separate COVID regressors	6.3 (0.3)	0.79	9.6 (0.8)	1.30
	Binary (shared)	7.1 (0.5)	0.92	6.3 (0.4)	0.88
	Raw stringency (shared)	30.4 (1.3)	3.76	34.8 (1.3)	4.75
	HESI (shared)	6.2 (0.3)	0.78	8.4 (0.7)	1.14
Prophet monthly	No calendar regressors	9.4 (0.7)	1.20	7.3 (0.4)	1.01
	COVID ignored	8.8 (0.7)	1.13	7.4 (0.5)	1.02
	Mask COVID window	6.3 (0.4)	0.78	6.3 (0.4)	0.87
	Separate COVID regressors	6.0 (0.3)	0.75	8.6 (0.7)	1.16
	Binary (shared)	9.3 (0.7)	1.20	6.9 (0.4)	0.95
	Raw stringency (shared)	23.6 (1.8)	2.92	34.3 (2.0)	4.66
	HESI (shared)	6.1 (0.3)	0.76	8.6 (0.7)	1.16
SARIMAX	COVID ignored	6.7 (0.3)	0.85	5.9 (0.3)	0.82
	HESI (shared)	**5.7 (0.3)**	0.71	**5.7 (0.3)**	0.79
LightGBM daily	HESI (shared)	13.4 (0.6)	1.72	8.7 (0.5)	1.26
LightGBM monthly	HESI (shared)	12.7 (0.6)	1.64	8.4 (0.5)	1.20

SARIMAX in S1 (0.199), the aggregation correction in S1 (0.070) and SARIMAX against masking in E2 (0.474 and 0.474); they should not be over-interpreted.

7 Discussion

7.1 What transfers, and what does not

The evidence supports the central hypothesis with qualifications. Elasticities identified on recurring holidays transfer to lockdowns when three conditions hold: the shock is expressed as a dose rather than a switch; the elasticity is estimated at a frequency at which holidays are identified, or the monthly estimate is corrected for aggregation; and the

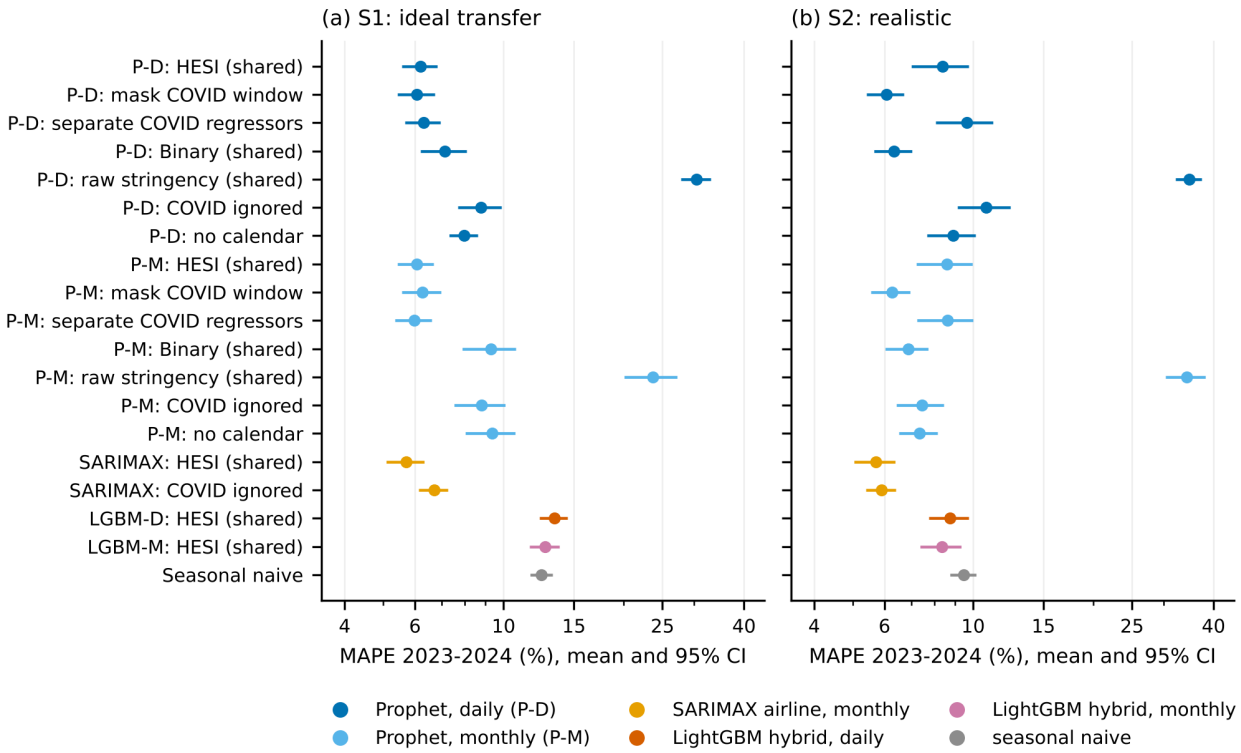


Figure 9. Aftermath forecasts (E2): mean MAPE for 2023–2024 with 95% confidence intervals across replications, logarithmic horizontal axis. Marker colours identify method families: blue for daily Prophet, sky blue for monthly Prophet, orange for SARIMAX, vermilion for daily LightGBM, lilac for monthly LightGBM and grey for the seasonal naive benchmark. Row labels give the treatment of the pandemic in the training sample.

Table 8. Normal period (E0, January 2018–December 2019): MAPE (%) with Monte Carlo standard errors and mean MASE.

Method	Specification	MAPE	MASE
Seasonal naive	–	7.5 (0.3)	1.08
Prophet daily	no calendar regressors	7.4 (0.6)	1.06
Prophet daily (P-D)	holiday regressors	7.3 (0.7)	1.05
Prophet monthly	no calendar regressors	7.2 (0.6)	1.03
Prophet monthly (P-M)	holiday regressors	6.9 (0.6)	0.99
P-M + AC	holiday regressors + AC	6.9 (0.6)	0.99
SARIMAX	holiday regressors	5.9 (0.4)	0.85
LightGBM daily	holiday features	**4.7 (0.2)**	0.67
LightGBM monthly	holiday features	5.5 (0.2)	0.79

dose scale is either known or calibrated on early data. When these conditions fail, the transfer degrades predictably: an overstated dose produces the right-hand branch of the U-curve, a monthly elasticity produces attenuation, and a tree ensemble produces saturation. The framework is also informative about its own failure: the calibrated dose absorbs coincident demand shocks (Proposition 1(b)), so a calibrated $\hat{\lambda}$ far from plausible pass-through values signals that the shock is not primarily a stoppage.

7.2 A protocol for official Thai export data

1. Use monthly merchandise exports excluding gold as the main series and record the data vintage.
2. Build the observed holiday calendar and the Chinese-New-Year windows.
3. Construct HESI from the Google Community Mobility Reports, with separate references for festivals and ordinary holidays (two-tier specification), and use episode indices or better proxies for China, never raw stringency.

Table 9. Paired comparisons across 30 replications. Δ is the mean difference in MAPE (method A minus method B, percentage points) with a 95% t-interval; negative values favour A. "A better" is the share of replications in which A has the lower MAPE. p-values are from the Wilcoxon signed-rank test, unadjusted and Holm-adjusted across all comparisons in the table.

Exp.	Comparison (A vs. B)	Scen.	Δ [95% CI]	A better	p	Holm p
E1, Jan 20–Dec 21	P-D HESI vs. P-D Off	S1	−18.2 [−20.8, −15.7]	100%	<0.001	<0.001
		S2	−3.9 [−7.1, −0.6]	70%	0.031	0.155
E1, Jan 20–Dec 21	P-D HESI vs. P-D Binary	S1	−14.9 [−16.9, −12.9]	100%	<0.001	<0.001
		S2	−4.7 [−7.4, −1.9]	67%	0.004	0.030
E1, Jan 20–Dec 21	P-D HESI vs. P-D stringency	S1	−19.5 [−22.2, −16.8]	97%	<0.001	<0.001
		S2	−23.5 [−24.2, −22.9]	100%	<0.001	<0.001
E1, May 20–Dec 21	P-D HESI-cal vs. P-D Off	S1	−17.2 [−20.5, −13.9]	100%	<0.001	<0.001
		S2	−4.2 [−7.1, −1.2]	63%	0.015	0.093
E1, May 20–Dec 21	P-D HESI-cal vs. P-D HESI	S1	−0.4 [−1.5, 0.7]	53%	0.556	0.556
		S2	−2.8 [−3.9, −1.7]	77%	<0.001	<0.001
E1, Jan 20–Dec 21	P-D HESI vs. P-M HESI	S1	−3.3 [−5.1, −1.5]	73%	<0.001	0.005
		S2	2.2 [1.0, 3.5]	23%	0.001	0.013
E1, Jan 20–Dec 21	P-D HESI vs. SARIMAX HESI	S1	−2.1 [−4.1, −0.1]	60%	0.050	0.199
		S2	2.4 [0.9, 3.9]	27%	0.002	0.014
E1, Jan 20–Dec 21	P-D HESI vs. LightGBM-D HESI	S1	−31.9 [−33.6, −30.2]	100%	<0.001	<0.001
		S2	−34.8 [−36.2, −33.3]	100%	<0.001	<0.001
E1, Jan 20–Dec 21	P-M+AC HESI vs. P-M HESI	S1	−1.3 [−2.7, 0.2]	83%	0.010	0.070
		S2	2.7 [1.3, 4.1]	23%	<0.001	0.005
E2, 2023–24	P-D mask vs. P-D ignored	S1	−2.7 [−3.9, −1.5]	77%	<0.001	0.001
		S2	−4.7 [−6.3, −3.1]	87%	<0.001	<0.001
E2, 2023–24	P-D HESI vs. P-D ignored	S1	−2.6 [−3.4, −1.7]	73%	<0.001	<0.001
		S2	−2.4 [−2.9, −1.8]	90%	<0.001	<0.001
E2, 2023–24	SARIMAX HESI vs. P-D mask	S1	−0.4 [−1.0, 0.3]	63%	0.198	0.474
		S2	−0.4 [−1.1, 0.4]	60%	0.158	0.474
E0, 2018–19	LightGBM-D Off vs. P-D Off	S1	−2.6 [−4.0, −1.3]	70%	<0.001	0.006

4. With daily or weekly customs data, use daily Prophet with shared regressors; otherwise use monthly Prophet with the aggregation correction alongside SARIMAX.
5. Calibrate the dose on months whose figures have been published, and report the uncertainty of $\hat{\lambda}$.
6. For the aftermath, compare masking, HESI and SARIMAX by rolling-origin evaluation (Tashman, 2000), with Diebold–Mariano tests and the small-sample correction of Harvey et al. (1997).
7. Report scenario-conditional forecasts separately from ex-ante forecasts.

7.3 Threats to validity

Internal validity. In S1 the data-generating process satisfies the assumptions of the framework by construction, which favours HES. Scenario S2, the robustness mechanisms and the dose sweep relax these assumptions in controlled ways, but other departures are possible. *Construct validity.* Workplace mobility measures attendance, not production; the holiday reference is a modelling choice; the Chinese episodes are stylised; and MAPE emphasises relative errors in low-value months. *External validity.* The export series is simulated and its parameters are calibrated rather than estimated; official exports also contain gold, price and exchange-rate movements; and the study covers one economy. *Statistical conclusion validity.* With 30 replications, Monte Carlo standard errors are small relative to the main effects but not to all secondary contrasts, which Table 9 makes explicit. Prediction intervals were not evaluated.

7.4 Directions for future research

The most important next step is the application to official Thai export statistics and, where available, to daily customs records. Further directions include festival-specific dose references; a Bayesian prior on the pass-through κ that is updated as data are published; hierarchical models by product and destination; hybrid estimators in which a tree ensemble is trained on the residual after removing the shared stoppage effect, combining the transfer of HES with the accuracy of trees in normal times; and applications to other economies with long festivals, such as the Vietnamese Tet.

8 Conclusion

Treating lockdowns as partial holidays turns an unforecastable shock into a measurable dose of a familiar event. When the work-stoppage elasticity is identified on recurring holidays through a shared regressor and the pandemic is expressed in holiday units, a forecasting model can anticipate the effect of a lockdown without observing one. The

theory developed here explains when this works and how it fails: through the dose scale, through temporal aggregation and through the piecewise-constant structure of tree ensembles. The simulation evidence confirms the theory quantitatively. The approach is not a substitute for adaptive models in normal times or after a shock has passed; it is a structured way to use the knowledge contained in every holiday of the past when the economy stops in an unprecedented way.

Data and code availability

The code is distributed with this preprint as ancillary files. The file `hesi_forecast.py` reproduces all results with `python hesi_forecast.py simulate -reps 30`; `robust_en.py` runs the robustness experiments and `figs_en.py` the figures, including an automated check that no text in any figure overlaps other text or data areas. The command `python hesi_forecast.py real -exports exports.csv` applies the monthly methods to official data (columns `month,value`), downloading OxCGRT and Google Mobility data automatically. All numbers in the text and tables are generated from the result files.

References

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47–78.

Baldwin, R., & Tomiura, E. (2020). Thinking ahead about the trade impact of COVID-19. In R. Baldwin & B. Weder di Mauro (Eds.), *Economics in the Time of COVID-19* (pp. 59–71). CEPR Press.

Bell, W. R., & Hillmer, S. C. (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association*, 78(383), 526–534.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349), 70–79.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.

Chen, C., & Liu, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421), 284–297.

Chow, G. C., & Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *Review of Economics and Statistics*, 53(4), 372–375.

Clements, M. P., & Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge University Press.

Cleveland, W. P., & Devlin, S. J. (1982). Calendar effects in monthly time series: Modeling and adjustment. *Journal of the American Statistical Association*, 77(379), 520–528.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, 66(333), 99–102.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383), 605–610.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2), 127–152.

Foroni, C., Marcellino, M., & Stevanovic, D. (2022). Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*, 38(2), 596–612.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), 53–90.

Google LLC. (2022). *Google COVID-19 Community Mobility Reports*. https://www.google.com/covid19/mobility/

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems 35, Datasets and Benchmarks Track*.

Hale, T., Angrist, N., Goldszmidt, R., et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5, 529–538.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.

Januschowski, T., Wang, Y., Torkkola, K., Erkkilä, T., Hasson, H., & Gasthaus, J. (2022). Forecasting with trees. *International Journal of Forecasting*, 38(4), 1473–1481.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30*.

Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after March 2020. *Journal of Applied Econometrics*, 37(4), 688–699.

Lin, J.-L., & Liu, T.-S. (2003). Modeling lunar calendar holiday effects in Taiwan. *Taiwan Economic Forecast and Policy*, 33(2), 1–37.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364.

Meta Open Source. (n.d.). Handling shocks. *Prophet documentation*. https://facebook.github.io/prophet/docs/handling_shocks.html

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361–1401.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1), 134–161.

python-holidays contributors. (2025). *holidays: Open World Holidays Framework* [Software]. https://github.com/vacanza/holidays

Rossi, B. (2013). Advances in forecasting under instability. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2B, pp. 1203–1324). Elsevier.

Sampi, J., & Jooste, C. (2020). Nowcasting economic activity in times of COVID-19: An approximation from the Google Community Mobility Report. *World Bank Policy Research Working Paper*, 9247.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96).

Silvestrini, A., & Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, 22(3), 458–497.

Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1), 11–30.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437–450.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.

A Proofs

Remark 1. Prophet represents holiday effects as $Z(t)\boldsymbol{\kappa}$ with $\kappa_j \sim N(0,\sigma_h^2)$. For a holiday that does not occur in the estimation window, column j of Z is zero for all estimation dates, so the log posterior equals $-\kappa_j^2/(2\sigma_h^2)$ plus terms that do not involve κ_j. It is maximised at $\kappa_j = 0$, and the forecast contribution $Z_j(t)\kappa_j$ is zero for every future date. □

Proposition 1. On a non-holiday day with $\lambda h_d \le 1$ and $z_d \le 1$, equation (1) gives $\log y_d = m_d + \beta_D \kappa h_d(1+\nu_d) + \beta_P x_d^P + D_d + u_d$ with $m_d = g + s + w$, and the forecast is $\log \hat{y}_d = \hat{m}_d + \hat{\beta}_D \lambda h_d + \hat{\beta}_P \hat{x}_d^P$. Subtracting and adding $\beta_D \lambda h_d$ yields (7) with $r_d = (m_d - \hat{m}_d) + (\beta_P x_d^P - \hat{\beta}_P \hat{x}_d^P) + D_d + u_d$. Part (a) follows by taking expectations with $\mathbb{E}[\nu_d \mid h_d] = 0$. For part (b), with $\hat{\beta}_D = \beta_D$ and ν_d independent of r_d, $\sum_d \mathbb{E}[e_d^2] = \sum_d \mathbb{E}[\{\beta_D h_d(\kappa - \lambda) + r_d\}^2] + \text{const}$. Setting the derivative with respect to λ to zero gives $\beta_D^2(\kappa - \lambda)\sum_d h_d^2 + \beta_D \sum_d h_d \mathbb{E}[r_d] = 0$, which is (8); the objective is convex in λ. Part (c) substitutes $\lambda h_d \equiv 1$ in the selected months and $\lambda = 0$ elsewhere. □

Proposition 2. Let X be a day drawn from month m with probabilities w_d. Then $E_m(\beta) = K(\beta)$, the cumulant generating function of X. (a) K is convex with $K(0) = 0$ and $K'(0) = \mathbb{E}_w X = \bar{x}_m$, so $K(\beta) \ge \beta \bar{x}_m$ (equivalently, Jensen's inequality applied to $e^{\beta X}$), with equality if and only if X is degenerate. (b) is the cumulant expansion $K(\beta) = \beta\kappa_1 + \beta^2\kappa_2/2 + O(\beta^3)$ with $\kappa_1 = \bar{x}_m$ and $\kappa_2 = \mathrm{Var}_w(X)$. (c) If $X \in \{0,1\}$ with $\Pr(X=1) = q$, then $K(\beta) = \log(1 - q + qe^\beta)$. Since $\log(1+u) = u + O(u^2)$, $K(\beta) = q(e^\beta - 1) + O(q^2)$, and the least-squares slope through the origin, $\sum qK / \sum q^2$, equals $e^\beta - 1 + O(\max q)$. Because $e^\beta > 1 + \beta$ for $\beta \ne 0$, $e^\beta - 1 > \beta$, so $|\beta_M| < |\beta|$ when $\beta < 0$. (d) A constant dose gives $K(\beta) = \beta c$, and the error of the monthly forecast is $(e^\beta - 1)c - \beta c = (e^\beta - 1 - \beta)c \ge 0$, i.e. the forecast effect is smaller in magnitude by $(1 + \beta - e^\beta)|c|$ for $c \ge 0$, since $\beta c \le 0$. □

Corollary 1. The map $\beta \mapsto e^\beta - 1$ is strictly increasing with inverse $b \mapsto \log(1+b)$ on $b > -1$. The delta method gives $\mathrm{Var}\{\log(1+\hat{\beta}_M)\} \approx \mathrm{Var}(\hat{\beta}_M)/(1+\beta_M)^2$. The minimum-distance version replaces the approximation $q(e^b - 1)$ by the exact $E_m(b)$ for each month. □

Proposition 3. The Frisch–Waugh–Lovell theorem expresses $\hat{\beta}_M$ as the slope of the regression of residualised $\log(Y_m/D_m)$ on residualised $\tilde{x}_m^D$; under homoskedastic, uncorrelated errors its variance is $\sigma_m^2/\sum_m \tilde{x}_m^2$. Holidays with fixed dates contribute the same amount to $\bar{x}_m^D$ in every year apart from in-lieu adjustments, so their contribution is absorbed by the calendar-month effects and does not enter $\tilde{x}_m$. In daily data the residual of a holiday indicator after projection on smooth seasonal and day-of-week terms is close to the indicator itself, because a single day has negligible projection on smooth functions of the date. □

Proposition 4. A regression tree splits a feature at thresholds between distinct observed values. With observed values $\{0,1\}$, every threshold lies in $(0,1)$; with histogram binning there is a single bin boundary and hence a single threshold c for all trees. Holding the other features fixed, every tree maps $x \le c$ and $x > c$ to two constants, and so does their sum. For a continuous feature with maximal observed value a, no threshold exceeds a, so all $x > a$ fall in the same leaves as $x = a$. □

B Additional results

Table 10. Shock period (E1): mean MASE over 30 replications with Monte Carlo standard errors.

		S1: ideal transfer		S2: realistic	
Method	Dose variant	Jan–Dec	May–Dec	Jan–Dec	May–Dec
Seasonal naive	–	2.37 (0.08)	2.34 (0.09)	1.44 (0.06)	1.34 (0.07)
Prophet daily	No calendar regressors	3.07 (0.15)	3.14 (0.17)	1.91 (0.11)	1.85 (0.13)
Prophet daily (P-D)	Off (no dose)	3.00 (0.17)	2.95 (0.19)	1.90 (0.12)	1.74 (0.14)
	Binary whole-month switch	2.69 (0.14)	2.85 (0.16)	2.06 (0.09)	2.05 (0.11)
	Raw stringency	3.43 (0.14)	3.88 (0.16)	4.83 (0.15)	5.44 (0.18)
	HESI ($\lambda = 1$)	**0.94 (0.06)**	0.99 (0.07)	1.55 (0.11)	1.69 (0.14)
	HESI-cal ($\hat{\lambda}$)	–	**0.95 (0.06)**	–	1.27 (0.10)
	Oracle (κ known)	0.94 (0.06)	0.99 (0.07)	1.15 (0.07)	1.21 (0.09)
Prophet monthly (P-M)	Off (no dose)	3.02 (0.14)	2.98 (0.16)	1.86 (0.11)	1.69 (0.12)
	Binary whole-month switch	2.52 (0.12)	2.71 (0.14)	1.76 (0.08)	1.78 (0.10)
	Raw stringency	2.19 (0.15)	2.46 (0.17)	3.46 (0.19)	3.92 (0.22)
	HESI ($\lambda = 1$)	1.28 (0.12)	1.31 (0.14)	1.20 (0.08)	1.28 (0.09)
	HESI-cal ($\hat{\lambda}$)	–	1.09 (0.08)	–	1.30 (0.11)
	Oracle (κ known)	1.28 (0.12)	1.31 (0.14)	1.23 (0.08)	1.23 (0.09)
P-M + AC	Binary whole-month switch	2.70 (0.11)	2.86 (0.13)	1.98 (0.09)	1.96 (0.10)
	Raw stringency	3.26 (0.23)	3.67 (0.26)	4.62 (0.25)	5.20 (0.28)
	HESI ($\lambda = 1$)	1.16 (0.11)	1.21 (0.13)	1.59 (0.13)	1.72 (0.15)
	HESI-cal ($\hat{\lambda}$)	–	0.97 (0.06)	–	1.24 (0.08)
	Oracle (κ known)	1.16 (0.11)	1.21 (0.13)	1.21 (0.08)	1.25 (0.09)
SARIMAX	Off (no dose)	2.87 (0.15)	2.82 (0.17)	1.72 (0.12)	1.55 (0.13)
	Binary whole-month switch	2.38 (0.13)	2.56 (0.15)	1.65 (0.09)	1.66 (0.11)
	Raw stringency	2.28 (0.15)	2.59 (0.17)	3.58 (0.18)	4.08 (0.20)
	HESI ($\lambda = 1$)	1.16 (0.12)	1.20 (0.14)	**1.19 (0.07)**	1.31 (0.08)
	HESI-cal ($\hat{\lambda}$)	–	1.05 (0.09)	–	1.25 (0.09)
	Oracle (κ known)	1.16 (0.12)	1.20 (0.14)	1.14 (0.08)	1.16 (0.10)
LightGBM daily	Off (no dose)	2.50 (0.10)	2.40 (0.10)	1.43 (0.07)	1.23 (0.07)
	Binary whole-month switch	2.30 (0.07)	2.38 (0.09)	1.70 (0.05)	1.62 (0.06)
	Raw stringency	5.07 (0.11)	5.62 (0.12)	6.47 (0.13)	7.18 (0.14)
	HESI ($\lambda = 1$)	4.97 (0.10)	5.45 (0.11)	6.37 (0.12)	7.02 (0.14)
	HESI-cal ($\hat{\lambda}$)	–	2.69 (0.20)	–	1.23 (0.07)
	Oracle (κ known)	4.97 (0.10)	5.45 (0.11)	6.37 (0.12)	7.02 (0.14)
LightGBM monthly	Off (no dose)	2.49 (0.10)	2.34 (0.11)	1.48 (0.06)	1.24 (0.07)
	Binary whole-month switch	2.31 (0.09)	2.21 (0.10)	1.33 (0.06)	1.14 (0.06)
	Raw stringency	1.77 (0.08)	1.69 (0.09)	1.20 (0.05)	**1.11 (0.05)**
	HESI ($\lambda = 1$)	1.90 (0.08)	1.83 (0.09)	1.22 (0.05)	1.12 (0.05)
	HESI-cal ($\hat{\lambda}$)	–	1.88 (0.10)	–	1.17 (0.05)
	Oracle (κ known)	1.90 (0.08)	1.83 (0.09)	1.23 (0.05)	1.12 (0.05)

C Implementation details

All computations used Python 3.12 with prophet 1.5.0, lightgbm 4.7.0, statsmodels 0.15.0 (Seabold and Perktold, 2010) and holidays 0.108. Prophet was estimated by maximum a posteriori with default trend priors, regressor prior scale 10 and `uncertainty_samples=0`. LightGBM used 400 trees, learning rate 0.05, 31 leaves and 20 observations per

leaf for daily data, and 300 trees, learning rate 0.05, 7 leaves and 5 observations per leaf for monthly data. SARIMAX was estimated by maximum likelihood with up to 300 iterations. One replication of all experiments takes about 25–30 seconds on one CPU core; the robustness experiments take about 2–3 seconds per replication.

D Code excerpt

```python
def mobility_hesi(mob, hol, window=7):
    """HESI_d = workplace-mobility drop on day d / median drop on weekday public holidays."""
    drop = (-mob / 100.0).clip(lower=0)
    is_hol = drop.index.isin(hol)
    ref = float(drop[is_hol & (drop.index.dayofweek < 5)].median())
    smooth = drop.where(~is_hol).rolling(window, center=True, min_periods=3).mean()
    return (smooth / ref).clip(0, 1).fillna(0.0), ref

def stoppage_X(cal, drv, variant, lam=1.0):
    """Shared regressors: xD = H + (1-H)*min(1, lam*HESI); xP = C + (1-C)*episode."""
    H, C = cal.H.values, cal.C.values
    zD, zP = np.zeros(len(cal)), np.zeros(len(cal))
    if variant == "binary": zD, zP = drv.bin_th.values, drv.bin_cn.values
    elif variant == "stringency": zD, zP = drv.s_th.values, drv.s_cn.values
    elif variant == "hesi": zD, zP = lam * drv.hesi_th.values, drv.e_cn.values
    zD, zP = np.clip(zD, 0, 1), np.clip(zP, 0, 1)
    return pd.DataFrame({"xD": H + (1 - H) * zD, "xP": C + (1 - C) * zP}, index=cal.index)

def deaggregate_beta(beta_m, x_train, grid=np.linspace(-3.0, 0.0, 601)):
    """Aggregation correction: daily beta whose monthly log-mean-exp matches beta_m * mean(x)."""
    Mx = month_matrix(x_train) # months x days (NaN padded)
    target = beta_m * np.nanmean(Mx, axis=1)
    E = np.log(np.nanmean(np.exp(grid[:, None, None] * Mx[None]), axis=2))
    return float(grid[np.argmin(((E - target[None]) ** 2).sum(axis=1))])
```