Traffic Congestion and Urban Heat: Evidence from Freeway Incidents in California

Sierra Arnold*

June 17, 2026

Abstract

Traffic congestion is a plausible but unquantified contributor to localized urban heat—a costly externality for the many who live and work near major roads—yet its short-run causal effect on local surface temperature has not been credibly identified; existing estimates are associational or rely on broad natural experiments. We estimate it for California freeways using unplanned accidents as an instrument for congestion. The central threat is reverse causality—congestion itself raises accident risk—which we address by restricting the instrument to accidents that strike while the segment is free-flowing, so congestion cannot have caused them; an event study validates the restriction. We measure land surface temperature from NASA ECOSTRESS at 70-meter resolution and exploit its varying overpass time to match each reading to same-hour congestion. The clean instrument is right-signed at every distance from the road. Our central estimate is the reduced form: a free-flow-onset accident raises near-road surface temperature by about 0.07°C over the following two hours, with flat pre-trends, significant at the 10 percent level under weak-instrument-robust inference. Scaled by the accident's effect on congestion, this implies roughly 0.3 to 0.7°C per 10 mph of speed reduction—a range whose bounds reflect how much of the first stage is attributed to the accident shock itself, and which brackets the prior associational estimate. The effect decays with distance and is statistically resolved only near the road, leaving the temperature response in surrounding residential areas suggestive rather than conclusive.

*Assistant Professor of Economics, Xavier University. Contact: arnolds4@xavier.edu. I thank participants at the ASHEcon Annual Conference and at the Miami University brown bag seminar for helpful comments and suggestions. Any errors are my own.

Keywords: urban heat island, traffic congestion, land surface temperature, instrumental variables, freeway incidents, ECOSTRESS, California

JEL Classification: Q53, R41, R12

1 Introduction

How far does the heat generated by traffic reach into the surrounding city? Extreme heat is the leading cause of weather-related deaths in the United States, exceeding floods, tornadoes, and hurricanes in average annual fatalities (National Weather Service, 2024). Urban heat—the systematic warmth of developed areas relative to their surroundings—carries well-documented health and energy costs (Heaviside et al., 2017; Santamouris et al., 2015), and its *structural* drivers, the pavement, rooftops, and sparse vegetation that set a neighborhood's baseline temperature, are extensively studied (Oke, 1973; Gago et al., 2013). Much less is known about its *time-varying* drivers, and in particular about traffic congestion, which fluctuates hour to hour over a fixed road network that millions of people live beside: about 11 million Californians reside within one kilometer of the freeway segments we study, 4.7 million of them within 500 meters. Whether congestion measurably heats these neighborhoods, and how far that heat reaches, is an understudied spatial question.

Vehicles in stop-and-go operation burn more fuel and release more waste heat per mile than vehicles at cruise speed (Barth and Boriboonsomsin, 2008), and congestion concentrates idling engines over a fixed road segment. The anthropogenic heat-flux literature identifies vehicular traffic as a distinct and substantial component of the urban energy balance, peaked at commute hours (Sailor and Lu, 2004; Sailor, 2011; Quah and Roth, 2012), and anthropogenic heat measurably raises urban temperatures (Ichinose et al., 1999). When congestion raises nearby surface temperatures, the exposure is concentrated in exactly the dense corridors where many people live and work (Wilson, 2020).

Causal identification of the congestion-to-heat link is limited. Lee and Berkelhammer (2025) correlate Chicago Transit Authority bus speeds with GOES satellite land surface temperature and estimates a 0.36°C increase in surface urban heat island intensity per 10 mph speed reduction; the estimate is associational, and congestion-prone corridors differ from uncongested ones in pavement, land use, and vegetation in ways that independently affect surface temperature. Quasi-experimental alternatives have relied on shocks, including

highway expansions (Yang et al., 2026) and COVID-19 mobility restrictions (Tran et al., 2026), that move land use, economic activity, and human presence alongside traffic.

We estimate the effect of freeway congestion on nearby surface temperature using unplanned freeway accidents logged in the Caltrans Performance Measurement System (PeMS) as an instrument for congestion. Accidents reduce effective road capacity and produce sudden, localized congestion on the affected segment. The central threat to this design is *reverse causality*: congestion itself raises accident risk, so the timing of an accident is not, in general, exogenous to congestion. We address this by restricting the instrument to *free-flow-onset* accidents, which are those that strike while the segment is free-flowing and therefore cannot have been caused by congestion.

We measure the temperature response with the NASA ECOSTRESS instrument (Hulley and Hook, 2022), which records land surface temperature (LST) at 70-meter resolution—far finer than the one-kilometer MODIS data used in prior work (Chakraborty et al., 2019)—and, because it orbits aboard the International Space Station rather than on a sun-synchronous track, overpasses each location at varying times of day. This lets us match each LST observation to a PeMS congestion reading from the same hour, a link unavailable with conventional satellites. We construct the outcome in concentric distance bands around the freeway: inner bands (100 and 250 meters) capture near-road temperature, while outer bands (500 meters and one kilometer) capture temperature in the surrounding residential area. The banded outcome lets a single design speak both to the temperature of the road and to the temperature experienced by nearby households—the population whose exposure is of most direct welfare interest, in the spirit of Currie and Walker (2011).

An event study is central to the design, and we report it before the instrumental-variables estimates (Section 4.3). It shows that congestion is flat—indeed slightly above free-flow—in the hours before a free-flow-onset accident and drops sharply at impact, whereas it ramps steeply upward before a congested-onset accident. This isolates the reverse causality in the congested-onset group and shows it is absent from the free-flow-onset group. Surface-

temperature pre-trends are flat for both groups, consistent with the exclusion restriction.

We report the estimates as a progression rather than a single headline number. Within detector-by-date-by-hour variation, the OLS association is small and wrong-signed, indicating that cross-sectional confounding dominates the raw correlation. The naive instrument using all accidents yields a precise, correctly signed estimate, but the event study shows it is contaminated by reverse causality, so we do not treat it as preferred. Our central estimate is the reduced form of the clean natural experiment: a free-flow-onset accident raises near-road surface temperature by about 0.07°C over the following two hours, with flat pre-trends, significant at the 10 percent level under weak-instrument-robust inference. Expressed per unit of congestion through two-stage least squares, this implies roughly 0.3 to 0.7°C per 10 mph reduction in speed. The bounds correspond to two defensible accountings of how much of the instrument's association with congestion is the accident shock itself; the range brackets the associational estimate in Lee and Berkelhammer (2025) (0.36°C per 10 mph). The effect is correctly signed at every distance band but is statistically resolved only near the road, so the temperature response in nearby residential areas is suggestive rather than conclusive.

The paper makes three contributions. First, it provides a credibly identified estimate of the short-run effect of traffic congestion on local surface temperature—a transparent natural-experiment reduced form with an instrumental-variables scaling—where prior evidence is associational. Second, it identifies reverse causality as a first-order problem for accident-based instruments and offers a transparent solution—restricting to free-flow-onset accidents—validated by an event study. Third, it uses the time-of-day variation and 70-meter resolution of ECOSTRESS to measure the temperature response in distance bands around the freeway, separating the on-road effect from the temperature experienced by nearby households.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature; Section 3 describes the data; Section 4 presents the research design; Section 4.3 presents the event study; Section 5 reports the results; Section 6 reports the robustness checks central

to identification, with further checks in Appendix; and Section 7 concludes.

2 Related Literature

This paper connects several bodies of work: the urban heat island literature, a small literature on traffic congestion and local temperature, and the causal economics literature on traffic externalities.

Urban heat islands arise when vegetated, permeable surfaces are replaced with impervious materials that absorb and retain solar radiation while eliminating evaporative cooling (Oke, 1973; Gago et al., 2013; Declet-Barreto et al., 2013). This literature centers on structural features of the built environment that vary slowly over time, measured at scale through thermal infrared remote sensing (Li et al., 2020); we study a time-varying driver, congestion, using the 70-meter, varying-overpass-time ECOSTRESS instrument (Fisher et al., 2020) to match temperature to traffic conditions hour by hour.

Direct evidence linking congestion to local temperature is scarce. Husni et al. (2022) document a 7°C temperature increase associated with traffic queue rates using IoT sensors in Indonesia; Yang et al. (2026) find that Bay Area highway expansions exacerbate urban–rural temperature differentials in MODIS data; and Lee and Berkelhammer (2025) estimate a 0.36°C increase in surface urban heat island intensity per 10 mph decline in Chicago bus speeds. These designs are associational; our contribution is causal identification of the same relationship, at finer spatial resolution and statewide scale.

A larger economics literature establishes that traffic imposes substantial external costs. These include infant health, productivity, and housing values (Greenstone and Gayer, 2018; Chay and Greenstone, 2003; Knittel et al., 2016; Xia et al., 2023; LSE Grantham Research Institute, 2022; Graff Zivin and Neidell, 2012; Chay and Greenstone, 2005). Currie and Walker (2011) use the introduction of E-ZPass, which reduced idling and congestion near toll plazas, as an instrument for local traffic conditions, finding improved birth outcomes

for nearby mothers. We extend this literature to the temperature channel, which operates through the same waste heat and exhaust as the pollution channel but along a thermal rather than chemical pathway. To instrument for congestion itself—rather than infrastructure, where Duranton and Turner (2011) established the canonical design—prior work has used transit strikes (LSE Grantham Research Institute, 2022) and weather (Knittel et al., 2016); unplanned freeway accidents offer finer spatial and temporal resolution than either, located to the detector segment and timestamped to the minute. This approach draws on a broad tradition in economics that exploits the idiosyncratic short-run timing of natural shocks—weather most prominently—as a source of exogenous variation (Miguel et al., 2004; Deschenês and Greenstone, 2007).

3 Data

3.1 Traffic Congestion and Incident Logs: PeMS

Traffic data come from the Caltrans Performance Measurement System (PeMS; California Department of Transportation, 2026), which operates a network of traffic detectors embedded in California freeway lanes. Each detector records flow (vehicles per hour), occupancy (share of time a vehicle is present over the sensor), and speed (miles per hour) at five-minute intervals. Detectors are spaced approximately every half mile along major California freeways and are organized into stations that aggregate across lanes. We obtain station-level hourly data and station metadata—including GPS coordinates, freeway route, and direction of travel—for eight Caltrans districts from June 2018 through March 2026.[1] Figure 1 maps the detectors in the analysis sample across these eight districts.

[1]We include the eight Caltrans districts covering California's major metropolitan areas: District 3 (Sacramento), District 4 (Bay Area), District 5 (San Luis Obispo/Central Coast), District 6 (Fresno/Central Valley), District 7 (Los Angeles), District 8 (San Bernardino/Riverside), District 11 (San Diego), and District 12 (Orange County). Districts 1 (North Coast), 2 (Northern California), 9 (Eastern Sierra), and 10 (Central Sierra) are excluded because they cover predominantly rural and mountainous terrain with sparse freeway detector networks and limited urban heat island exposure.

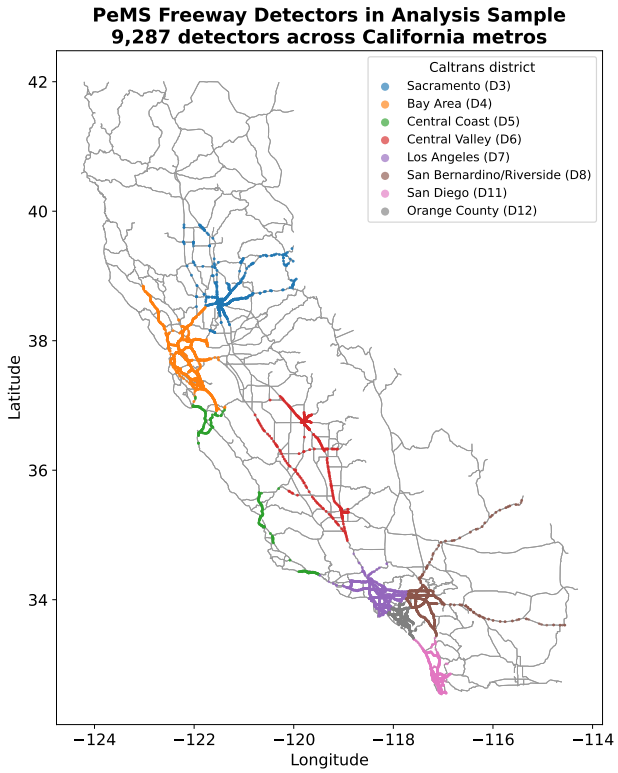

Figure 1: PeMS freeway detectors in the analysis sample, by Caltrans district. The sample spans the eight districts covering California's major metropolitan areas; detectors trace the instrumented freeway network in each metro.

The primary congestion measure is the *congestion index*, the ratio of realized speed to free-flow speed—a standard speed-based congestion metric also known as the speed performance index (He et al., 2016):

$$\text{CI}_{it} = \frac{\text{Speed}_{it}}{\text{Free-Flow Speed}_i}$$

where i indexes the PeMS detector station and t indexes the hour of the ECOSTRESS overpass. A value of 1 indicates free-flow conditions; values approaching 0 indicate near-standstill conditions. Free-flow speed $\overline{v}_i$ is defined empirically as the 85th percentile of observed speeds at station i during overnight hours (10pm–5am) and weekends, when congestion is minimal. This percentile-based definition makes the congestion index comparable across segments with different posted speed limits and traffic compositions.

For the incident instrument, we use PeMS incident logs, which record the timestamp, segment, type, and duration of each incident logged on California freeways. PeMS classifies incidents into a range of categories, the most common of which are traffic collisions, traffic hazards, animal hazards, vehicle fires, road and weather conditions, and assists to Caltrans operations. We restrict the instrument to traffic accidents, which are the most plausibly exogenous to neighborhood characteristics and environmental conditions. Weather-related incidents are excluded because precipitation simultaneously raises accident rates and suppresses LST, creating a mechanical negative correlation that would bias the instrument. Traffic hazards, animal hazards, vehicle fires and debris, and planned events such as construction assists and road closures are excluded because they either do not reliably generate congestion or are not unplanned. We discuss these restrictions further in Section 4.2.3.

The PeMS incident archive is largely complete over the study period, except for March-early July 2025, where the feed was not reliably archived.[2] We treat the instrument as undefined on dates lacking reliable incident records (missing, empty, or anomalously sparse) and *drop* ECOSTRESS overpasses falling in those windows. The 2025 gap therefore removes roughly four months of overpasses from the estimation sample. Because the gap is a data-archiving artifact unrelated to traffic, weather, or neighborhood conditions, this attrition is plausibly unrelated to the outcome and is not expected to bias the estimates; we report the affected sample counts in Table 1.

3.2 Land Surface Temperature: ECOSTRESS

Land surface temperature data come from the NASA ECOSTRESS `ECO_L2T_LSTE` Version 2 product, distributed by the NASA Land Processes Distributed Active Archive Center and accessed via AppEEARS (Hulley and Hook, 2022). ECOSTRESS is a thermal infrared instrument on the International Space Station (ISS) that has provided observations since July

[2]The remaining gaps are minor and isolated: a single day in June 2019, a cluster of days in January 2020, and scattered individual dates, together accounting for under two percent of days in the study period. Incident coverage in all other months is complete.

2018 at 70-meter spatial resolution (Fisher et al., 2020). The 70-meter resolution is substantially finer than the 1-kilometer MODIS data used in prior UHI research (Chakraborty et al., 2019) and allows LST measurement at the scale of individual freeway segments and adjacent land parcels.

The defining temporal property of ECOSTRESS for this study is its ISS orbit. Unlike sun-synchronous satellites that overpass at a fixed local solar time (Landsat 8 overpasses at approximately 10am; MODIS at fixed morning and afternoon times), the ISS precesses through a range of local overpass times at any given California location. This means successive ECOSTRESS overpasses of the same segment occur at different times of day, generating exogenous variation in the hour of LST observation; Figure 2 plots the distribution of overpass times across the 24-hour day. We exploit this variation to match each LST observation to a contemporaneous PeMS congestion reading at the overpass hour. The hour-of-day variation in overpass time also underlies the event study evidence described in Section 4.3.

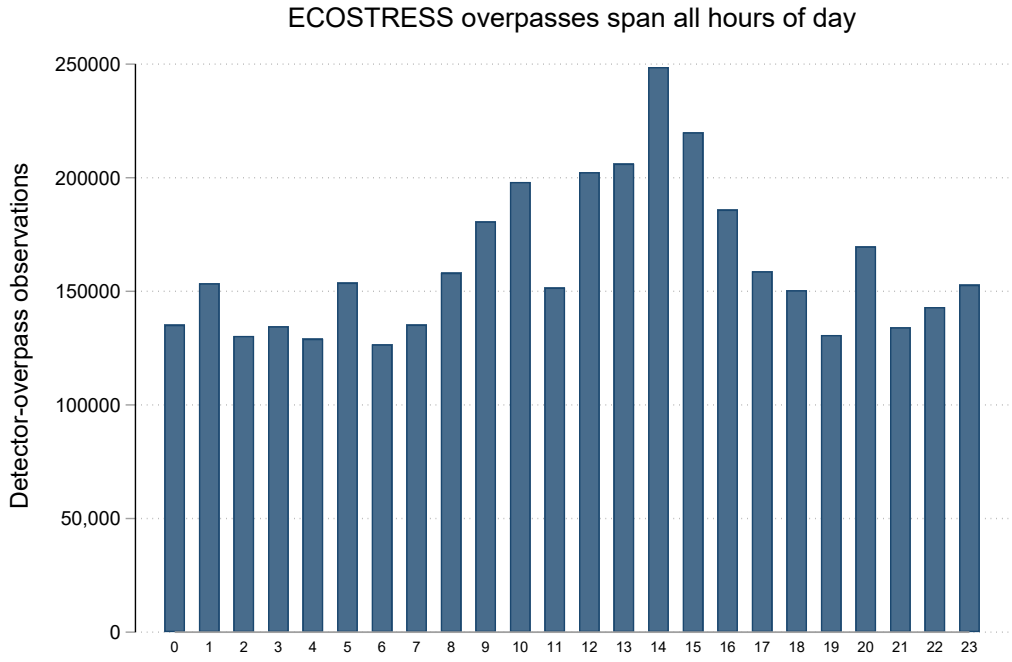

Figure 2: Distribution of ECOSTRESS overpass times by hour of day (Pacific) in the matched sample. Because the International Space Station orbit is not sun-synchronous, overpasses occur across all 24 hours rather than at a fixed local solar time, generating the time-of-day variation in land surface temperature that the identification strategy exploits.

We download four layers from ECO_L2T_LSTE: LST (in Kelvin), LST_err (retrieval uncertainty), QC (quality control flags), and cloud (cloud mask). We apply quality filters to exclude pixels where the LST retrieval is unreliable. Three sources of retrieval failure motivate these filters. First, clouds absorb and re-emit thermal radiation before it reaches the sensor, so a cloudy pixel records the temperature of the cloud rather than the ground surface; we exclude pixels flagged as cloudy by the cloud mask. Second, the retrieval algorithm corrects the raw thermal signal for atmospheric absorption and scattering, but this correction can fail under unusual water vapor or aerosol conditions— including, potentially, haze from heavily congested corridors—producing biased LST values; we exclude pixels with QC flag values indicating failed or degraded retrievals. Third, the algorithm converts radiance to temperature using assumed surface emissivity, and high uncertainty in that assumption produces inaccurate temperatures. We therefore drop pixels whose retrieved LST falls outside a physically plausible range (255–345 K, about −18 to 72°C), which removes implausible retrievals that survive the cloud and quality-control screens. The per-pixel retrieval-uncertainty layer (LST_err) is downloaded and available but is not applied as a screen in the baseline specification. Using corrupted pixels would add noise to the outcome variable and could introduce systematic bias if retrieval quality is correlated with congestion levels. After applying these filters, the outcome variable for observation (i, t) is the mean LST across all pixels passing quality filters within the 500-meter buffer of detector i at overpass t. The spatial footprint is restricted to a 2-kilometer buffer around the California state highway network, encompassing all PeMS detector locations and the immediately surrounding land.

Because ECOSTRESS is a relatively young instrument, coverage in the earliest months of our study period is incomplete. ECOSTRESS experienced two anomalies of its onboard mass storage unit (MSU) that suspended data acquisition: the primary unit failed on September 29, 2018 and acquisition resumed only after the instrument was switched to its backup unit on December 5, 2018, and a second anomaly halted acquisition again between March 14

and May 15, 2019 (NASA Land Processes Distributed Active Archive Center, 2024). Our panel therefore contains no LST observations for October through December 2018 and reduced coverage in the spring of 2019. Because these gaps arise from instrument hardware status rather than from any feature of traffic, weather, or location, the lost observations are unrelated to the outcome and affect only the early-period sample size, not the identifying variation.

3.3 Sample Construction

The unit of observation is a freeway detector station i crossed with an ECOSTRESS overpass t. A valid observation requires: (1) at least five ECOSTRESS pixels passing quality filters within the buffer of the detector; (2) a valid PeMS congestion reading within the same calendar hour as the overpass timestamp; and (3) that congestion reading passing the data-quality filter—at least 50 percent of the hour's five-minute intervals observed (`pct_observed` ≥ 50)—to reduce measurement error from detector malfunctions.

3.3.1 Outcome: distance bands around the freeway

The outcome is land surface temperature in concentric distance bands around the detector, each the mean of quality-filtered ECOSTRESS pixels within a given radius at overpass t. We construct the 100-, 250-, 500-, and 1,000-meter buffers, and from them three disjoint annular rings (0–100, 100–500, and 500–1,000 meters) for the distance gradient. The 100-meter band captures near-road temperature, where traffic-related heat is most concentrated: the epidemiological literature on near-road pollution—driven by the same combustion as vehicular waste heat—finds concentrations decaying to background within 150 to 300 meters and detectable to roughly 500 meters (Health Effects Institute, 2010). The outer bands capture temperature in the surrounding residential area, the exposure of most direct welfare interest, in the spirit of Currie and Walker (2011), who study residential outcomes for mothers within two kilometers of toll plazas. At 70-meter resolution a 500-meter radius contains about

160 pixels, yielding a stable average; we take the 500-meter buffer as the descriptive baseline, while the preferred causal estimate uses the near-road 100-meter band. Section 5 reports every band, and the gradient across the disjoint rings maps how far the effect reaches.

Several categories of observations are necessarily excluded from the analysis sample, and we report the resulting counts in Table 1. On the traffic side, not every PeMS detector in the monitored network contributes usable data: detectors that report no valid speed reading, that fall below 50 percent observation completeness in a given hour, or that came online after the start of the study period do not enter the panel for the affected periods. A small number of detectors that appear in the hourly traffic feed lack valid geographic coordinates in the station metadata; because these cannot be matched to ECOSTRESS pixels, they are dropped. On the temperature side, ECOSTRESS pixels failing the cloud, quality-control, or emissivity-uncertainty filters described in Section 3.2 are removed before aggregation, and a detector-overpass observation is retained only if at least five quality-passing pixels fall within its buffer, ensuring the LST average is not driven by one or two noisy pixels. Finally, an ECOSTRESS overpass enters the matched sample only if a valid PeMS congestion reading exists on the nearest detector within the same calendar hour; overpasses without a contemporaneous traffic reading are dropped. These restrictions reduce the sample to the set of detector-overpass pairs for which both a reliable congestion measure and a reliable surface temperature measure are available. Because detector reporting reliability and cloud cover are plausibly unrelated to the idiosyncratic timing of freeway incidents, we do not expect this attrition to threaten the identification strategy; we nonetheless report the share of observations lost at each filtering stage in Table 1 for transparency. Table 2 reports summary statistics for the resulting estimation sample.

Table 1: Sample construction

	Observations	Detectors	% of top
ECOSTRESS overpass observations	9,045,115	11,180	—
with ≥ 5 quality-passing pixels (500 m buffer)	8,910,223	11,178	98.5
matched to a same-hour PeMS congestion reading	4,124,564	9,323	45.6
excluding low-coverage incident days	3,890,544	9,287	43.0
Estimation panel	**3,890,544**	**9,287**	43.0
with non-missing weather controls	3,702,727	8,889	40.9

Notes: The unit of observation is a detector × ECOSTRESS overpass. Row 1 counts overpasses with a valid LST extraction (June 2018–March 2026; after pixel-level cloud, quality-control, and 255–345 K masking). The large drop at the same-hour PeMS match (row 3) reflects detector intermittency, not selection: the match rate is flat across all 24 hours of the day, so the attrition is orthogonal to the overpass-time variation the design exploits. Row 4 drops overpasses on dates with missing or anomalously sparse CHP incident records (the March–July 2025 feed gap and scattered days), which cannot be coded incident = 0. The final row is the sample for the preferred weather-controlled specification.

Table 2: Summary statistics

	Mean	SD	Min	Max	N
Land surface temp (C)	22.76	11.46	-18.05	71.43	3890544
Congestion index (speed / free-flow)	0.92	0.12	0.04	1.00	3890544
Speed (mph)	63.49	8.72	3.00	95.80	3890544
free_flow_speed	68.65	2.23	39.00	86.00	3890544
Traffic flow (veh/hr)	2674.12	2053.65	0.00	18509.00	3890544
Incident in prior 2h (0/1)	0.00	0.05	0.00	1.00	3890544
Peak-hour overpass (0/1)	0.25	0.43	0.00	1.00	3890544
pixel_count	142.50	35.53	5.00	164.00	3890544
Observations	3890544				

4 Research Design

4.1 Identification Problem

The parameter of interest is the causal effect of traffic congestion on land surface temperature. Let LST_{it} denote surface temperature at detector location i at overpass time t, and let CI_{it} denote the congestion index on the corresponding segment. The population relationship of interest is:

$$\text{LST}_{it} = \alpha_i + \mu_d + \lambda_h + \delta \cdot \text{CI}_{it} + \mathbf{X}'_{it}\gamma + u_{it}$$

where α_i are detector fixed effects, μ_d are date fixed effects (where d indexes the calendar date of overpass t), λ_h are hour-of-day fixed effects (where h indexes the Pacific-time hour of the overpass, $h \in \{0, 1, \ldots, 23\}$), $\mathbf{X}_{it}$ is a vector of local weather controls, and u_{it} is an error term. The coefficient δ is the causal effect of a one-unit decrease in the congestion index—corresponding to a shift from free-flow to fully congested conditions—on surface

temperature in degrees Celsius. We measure LST_{it} in concentric distance bands around the segment (Section 3.3.1); the baseline outcome is the mean over the 500-meter buffer, and Section 5 reports δ for each band, from the roadway out to one kilometer.

The choice of date fixed effects and hour-of-day fixed effects, rather than a single overpass-level time fixed effect, is deliberate. Detector fixed effects α_i absorb all time-invariant characteristics of the freeway location— pavement composition, land cover, urban density, and elevation—so that δ is identified from within-location deviations in LST. Date fixed effects μ_d absorb statewide weather shocks common to all segments on a given calendar day—heat waves, cloud cover, precipitation—without eliminating the within-day variation in overpass time. Hour-of-day fixed effects λ_h then absorb the systematic solar radiation cycle within a day, so that identification comes from deviations in congestion across overpasses of the same segment at the same time of day on different dates. The controls $\mathbf{X}_{it}$ include air temperature, precipitation, relative humidity, and wind speed at the detector location and overpass hour, absorbing local weather conditions that vary across segments and are not captured by the date fixed effects. Together, the fixed effects and controls mean that δ is identified from variation in congestion across overpasses of the same segment on the same calendar date and at the same time of day, purged of all time-invariant location characteristics and common daily shocks.

OLS estimation of δ is biased because CI_{it} is endogenous. Even within a given detector segment and time period, congestion may be correlated with time-varying unobservables that affect LST. Most importantly, clear-sky days have both higher solar radiation—which raises surface temperature through direct heating—and higher traffic volumes (fewer weather-related delays, more outdoor activity), which raises congestion. This clear-sky selection would generate a positive correlation between congestion and LST even in the absence of a causal effect, upward-biasing OLS. Conversely, hot afternoons may induce some drivers to avoid peak-hour travel, attenuating the OLS estimate. The direction and magnitude of the net bias is unclear *ex ante*, motivating the IV approach. The instrument we use—freeway

accidents—raises a distinct identification concern of its own, reverse causality, since congestion increases accident risk. We confront that directly in the instrument definition below and validate the solution with the event study (Section 4.3).

4.2 The Free-Flow-Onset Instrument

4.2.1 Definition

Our instrument for CI_{it} is an indicator for a *free-flow-onset* accident on segment i in the hours before the overpass, built in two steps. First, we flag any qualifying accident—a collision, hit-and-run collision, or spinout (Section 3.1)—on segment i within a lookback window of length h, with $h = 2$ hours in the baseline. Second, we keep only accidents that struck while the segment was free-flowing, defined as a congestion index of at least 0.90 in the hour before impact, so that congestion cannot have caused the accident:

$$Z_{it} = \mathbf{1}[\text{free-flow-onset accident on segment } i \text{ in } [t-h, t]]\,.$$

This restriction is the heart of the design. As the event study shows (Section 4.3), accidents in general are preceded by rising congestion, so an all-accident instrument would embed reverse causality, whereas free-flow-onset accidents are not and deliver a clean congestion shock. We report estimates using the all-accident instrument alongside the free-flow-onset instrument so that the contamination is visible; the free-flow-onset estimate is our preferred one.

The two-hour baseline lookback is the window over which the event study shows the congestion shock and its temperature response concentrated: the surface-temperature response appears in the $[0, 2)$-hour bin after impact (Section 4.3). We do not take a strong stand on the exact window and report $h \in \{1, 3\}$ hours as robustness (Appendix).

4.2.2 Relevance

For relevance, the instrument must predict congestion on the affected segment. Accidents reduce effective capacity, slowing traffic as vehicles merge around the incident; the resulting congestion can persist for one to three hours after clearance (Cao et al., 2021; Haule et al., 2019). Free-flow-onset accidents perturb congestion less than the average accident—they strike uncongested segments—so their first stage is smaller in magnitude, but it remains strong by the Olea and Pflueger (2013) benchmark of 16: the Kleibergen–Paap first-stage F is approximately 30–48 across the main specifications, although not so large that weak-instrument concerns can be dismissed outright, which is why we also report weak-instrument-robust inference (Section 5). We expect strength to rise with severity; Robustness 3 weights the instrument by accident duration, as longer accidents sustain larger congestion. A finer split by the number of lanes closed would help but is not available in the incident records.

4.2.3 Exclusion Restriction

The exclusion restriction requires that incident occurrence affects LST only through its effect on traffic congestion, conditional on the fixed effects and controls. We identify three potential violations and address each empirically; the event study in Section 4.3 provides the primary test, with the robustness checks in Section 6 providing supporting evidence.

Volume-driven clustering. Accident probability rises with traffic volume, which is correlated with congestion. Detector fixed effects α_i absorb the time-invariant component of this relationship; date and hour-of-day fixed effects absorb systematic volume patterns by calendar day and time of day. The residual question is whether within-segment, within-date accident timing correlates with concurrent temperature determinants beyond what the fixed effects and weather controls capture. We treat this as an empirical question rather than an assumption. The event study's pre-accident LST coefficients are the key test: flat pre-trends indicate that accident timing is not systematically related to unobserved temperature determinants at the time of the overpass. A prior-day placebo (Robustness 1) and weather

orthogonality check (Robustness 2) provide further evidence. We omit contemporaneous traffic volume from the main specification because it is jointly determined with congestion and is itself reduced by accidents—conditioning on it would constitute a bad control that absorbs part of the causal pathway (Angrist and Pischke, 2009). Robustness 11 (Appendix) confirms the estimates are insensitive to its inclusion.

Weather. Precipitation raises accident rates and lowers LST through evaporative cooling and reduced solar radiation. Date and hour-of-day fixed effects absorb most of this pattern; weather controls $\mathbf{X}_{it}$ address residual local variation. The weather orthogonality test (Robustness 2) assesses residual instrument-weather correlation after conditioning on the fixed effects.

Direct thermal effects and incident classification. Vehicle fires emit heat directly at the road surface, elevating pixel temperatures through a non-congestion channel. We exclude vehicle fires; qualifying incidents— collisions, hit-and-run collisions, and spinouts—generate no localized thermal signal at the 70-meter pixel scale, and the thermal footprint of emergency vehicles is negligible relative to the 500-meter buffer.

4.3 Event Study

The event study is the core of our identification, and we present it before the instrumental-variables estimates. It answers two questions. First, is the timing of an accident exogenous to congestion, or is congestion already building beforehand (reverse causality)? Second, are there pre-accident trends in surface temperature that would violate the exclusion restriction? The answers motivate restricting the instrument to free-flow-onset accidents.

4.3.1 Design

For each ECOSTRESS overpass we measure the elapsed time to each qualifying accident on the same detector within a ±6-hour window and assign the overpass to a two-hour lag bin

$[\ell, u)$:

$$d_{\ell,u,it} = \mathbf{1}[t - \text{accident}_{it} \in [\ell, u)] \,.$$

Negative lags are overpasses *before* an accident; the baseline group has no accident on the segment within ±6 hours. We estimate

$$Y_{it} = \alpha_i + \mu_d + \lambda_h + \sum_{[\ell,u)} \phi_{\ell,u}\, d_{\ell,u,it} + \eta_{it} \tag{1}$$

separately for $Y_{it} = \text{CI}_{it}$ (congestion index) and $Y_{it} = \text{LST}_{it}$ (degrees Celsius), with detector, date, and hour-of-day fixed effects and standard errors clustered at the detector. The lead coefficients $\phi_{\ell,u}$ at negative lags are the diagnostic of interest.

4.3.2 All accidents: congestion rises before impact

Estimated on all accidents, the instrument is contaminated. Congestion is not flat before impact: the lead coefficients rise steadily over the hours before an accident (-0.007, -0.023, -0.060) and peak at the accident bin (-0.076), roughly symmetric around impact—the signature of reverse causality, with accidents occurring disproportionately during transient congestion (Appendix Figure 6). Surface-temperature leads, by contrast, are flat and insignificant, with the only significant coefficient at the accident bin (+0.09°C; Appendix Figure 7): the exclusion restriction looks credible, but the all-accident instrument is not exogenous to congestion itself.

4.3.3 Free-flow-onset accidents: the clean variation

To separate exogenous from congestion-induced accidents, we classify each accident by the congestion index on its segment in the hour before impact. A *free-flow-onset* accident strikes while the segment is free-flowing (pre-accident CI ≥ 0.90) and so cannot have been caused by congestion; a *congested-onset* accident strikes while the segment is already congested. Among accidents we can classify, the two groups are roughly equal in number. We re-

estimate equation (1) with separate bin sets for the two groups in a single regression, so that congested-onset overpasses do not enter the free-flow-onset baseline (Figures 3–4).

The two groups behave very differently in congestion (Figure 3). Before a congested-onset accident, congestion ramps steeply (-0.024, -0.069, -0.151 across the three lead bins)—the reverse causality is concentrated here. Before a free-flow-onset accident, congestion does the opposite: the segment is, if anything, increasingly free-flowing ($+0.009$, $+0.022$, $+0.029$) and then drops sharply at impact (-0.014, $t \approx -7$). Free-flow-onset accidents thus deliver a clean, discontinuous congestion shock with no pre-trend. Surface-temperature leads are flat for both groups (Figure 4); for the free-flow-onset group the three lead coefficients are individually and jointly indistinguishable from zero ($F(3,\ 9{,}250) = 0.25$, $p = 0.86$), consistent with the exclusion restriction. The free-flow-onset accident bin shows a positive but imprecise temperature response.

These patterns motivate the rest of the analysis: we use free-flow-onset accidents as the instrument, because their congestion variation is not contaminated by reverse causality. The cost is that free-flow-onset accidents perturb congestion less than the average accident, which bears on precision (Section 5).

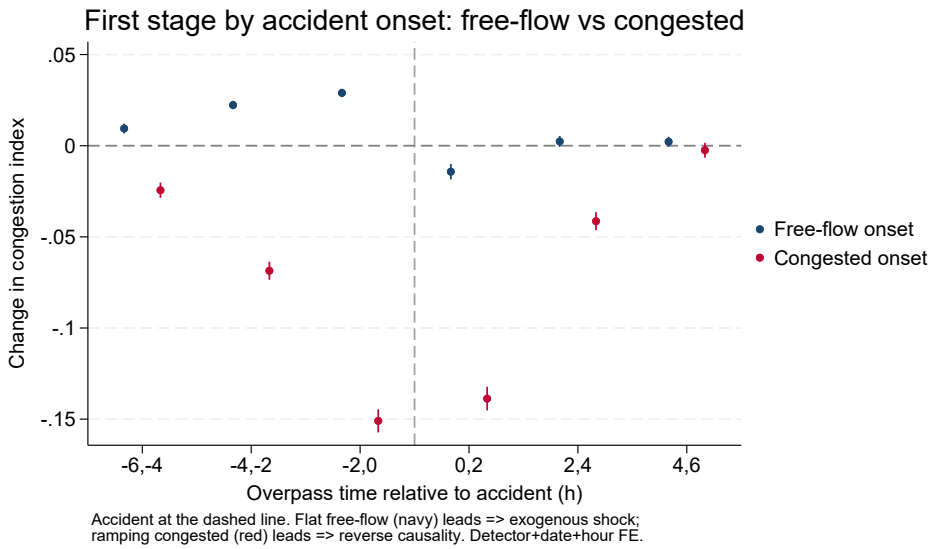

Figure 3: Event study by accident onset: congestion index. Free-flow-onset accidents (segment free-flowing before impact) versus congested-onset accidents, estimated jointly. Congestion ramps up before congested-onset accidents (reverse causality) but is flat—slightly above free-flow—before free-flow-onset accidents, dropping only at impact. Detector + date + hour-of-day FE; 95% CIs clustered at the detector.

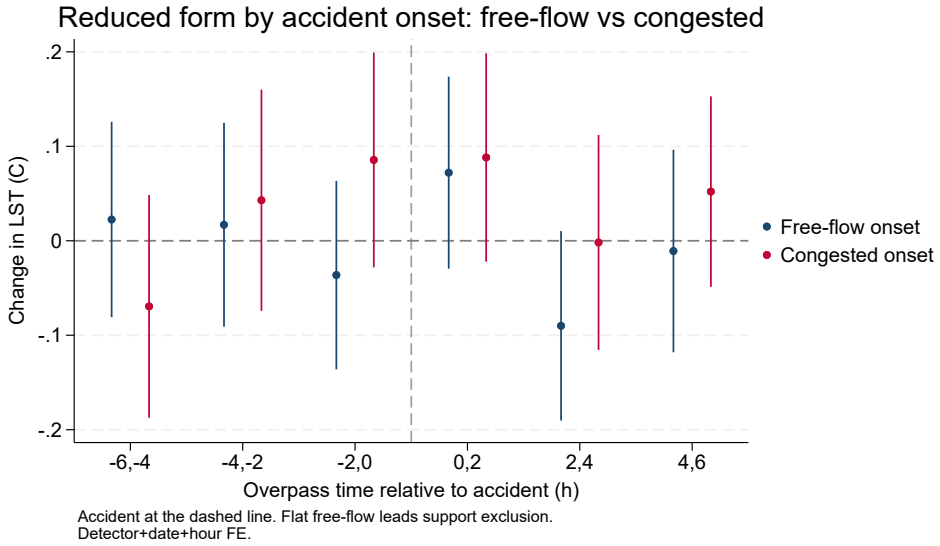

Figure 4: Event study by accident onset: surface temperature. Same specification as Figure 3, Y = LST (°C). Pre-accident leads are flat for both groups, consistent with the exclusion restriction.

4.4 Two-Stage Least Squares Specification

4.4.1 First Stage

The first stage regresses the congestion index on the free-flow-onset instrument Z_{it}, detector, date, and hour-of-day fixed effects, and weather controls:

$$\mathrm{CI}_{it} = \alpha_i + \mu_d + \lambda_h + \beta \cdot Z_{it} + \mathbf{X}'_{it}\gamma + \varepsilon_{it} \tag{2}$$

The coefficient β captures the average change in the congestion index associated with a free-flow-onset accident in the lookback window on segment i at overpass t. We expect $\beta < 0$: accidents reduce the speed ratio (increase congestion). The Kleibergen–Paap first-stage F-statistic on Z_{it} is the primary diagnostic for instrument strength.

4.4.2 Second Stage

The second stage replaces CI_{it} with its first-stage predicted value:

$$\mathrm{LST}_{it} = \alpha_i + \mu_d + \lambda_h + \delta \cdot \widehat{\mathrm{CI}}_{it} + \mathbf{X}'_{it}\gamma + u_{it} \tag{3}$$

The coefficient δ is the 2SLS estimate of the causal effect of a one-unit decrease in the congestion index on LST in degrees Celsius. Because the congestion index ranges from 0 (complete standstill) to 1 (free-flow), δ represents the temperature difference between the two extremes of the congestion distribution at a given location and time. We also report the implied effect of a 10 mph speed reduction—the unit used by Lee and Berkelhammer (2025)—to facilitate comparison with the existing literature. We estimate equation (3) for each distance band, so that δ traces the temperature response from the roadway out to the surrounding residential area (Section 5).

4.4.3 Fixed Effects and Controls

Detector fixed effects α_i control for all time-invariant characteristics of the freeway location and its immediate surroundings, including pavement composition, urban density, vegetation cover, and lane configuration. Date fixed effects μ_d control for all detector-invariant conditions on a given calendar day, including statewide weather patterns, any other common shocks, and the seasonal pattern of surface temperature. Hour-of-day fixed effects λ_h (one indicator per Pacific-time hour of the overpass, $h \in \{0, \ldots, 23\}$) absorb the systematic within-day variation in LST driven by the solar radiation cycle, ensuring that the instrument's predictive power is not contaminated by correlation between accident timing and the time of day. Together, these three-way fixed effects restrict identification to within-segment, within-date, within-hour deviations in LST and congestion.

Controls $\mathbf{X}_{it}$ include air temperature, precipitation, relative humidity, and wind speed at the detector location and overpass hour. These absorb local weather conditions within a day that vary across segments and are not fully captured by the date fixed effects. Air temperature and humidity affect LST through the atmospheric boundary layer; precipitation affects both the LST retrieval and incident occurrence; wind speed affects heat dissipation from the road surface.

4.4.4 Estimation

All specifications are estimated using `ivreghdfe` in Stata 18, which implements 2SLS with high-dimensional fixed effects (Baum et al., 2007). Standard errors are clustered at the detector-segment level to account for serial correlation in congestion and surface temperature within a segment across overpass times. Two-way clustering at the detector and date levels leaves the standard error on the preferred estimate essentially unchanged (SE 2.52 versus 2.52; $p = 0.033$), so the inference is not sensitive to common date-level shocks beyond what the date fixed effects already absorb. Because adjacent detectors can share ECOSTRESS pixels, we also re-cluster at progressively coarser spatial units—roughly 5- and 10-kilometer

grid cells and the freeway-by-direction corridor—which let nearby detectors correlate arbitrarily within a satellite scene (Table 3). The standard error rises only modestly, from 2.52 at the detector level to 2.73 at the corridor level; the estimate remains significant at five percent under grid-cell clustering and sits at the margin ($p = 0.05$) under the most conservative corridor clustering, so cross-detector spatial correlation does not materially understate the uncertainty.

Table 3: Spatial standard errors for the preferred estimate

	(1) Detector	(2) 5 km grid	(3) 10 km grid	(4) Corridor
Congestion index	-5.357**	-5.357**	-5.357**	-5.357*
	(2.518)	(2.494)	(2.579)	(2.733)
Clusters	8,852	945	464	194
Observations	3512527	3512527	3512527	3512527

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Because our first-stage F-statistics, though comfortably above the Olea and Pflueger (2013) threshold of 16, are not large in absolute terms, we do not rely on the conventional t-ratio alone for inference on the preferred estimate. Lee et al. (2022) show that for just-identified instrumental-variables estimation the usual critical value of 1.96 controls test size at the five percent level only when the first-stage F exceeds roughly 105; at smaller F their tF procedure calls for a larger critical value, so a conventional t-ratio overstates significance. At our first-stage strength ($F \approx 26$–30) this adjustment is material. We therefore report Anderson–Rubin weak-instrument-robust confidence sets for the preferred estimate (Section 5); these are valid regardless of instrument strength and place the near-road effect at the ten percent rather than the five percent margin. We regard this as the appropriate inference for what is, on the evidence, a borderline result.

A related caveat concerns the interpretation of the estimand rather than its precision. Blandhol et al. (2022) show that two-stage least squares with covariates recovers a convex, positively weighted average of heterogeneous treatment effects—the standard local-

average-treatment-effect interpretation—only when the covariates enter through a saturated specification. With the high-dimensional but non-saturated fixed effects we employ, the 2SLS estimand need not be a positively weighted average, so we treat δ as a local, design-based summary of the congestion–temperature response rather than a structural parameter, and we lean on the reduced-form event-study evidence (Section 4.3), which does not require this assumption, for the qualitative conclusion. We return to the local and extrapolation-sensitive character of the magnitude in Section 5.4.

5 Results

The estimation sample is the full panel of roughly 3.9 million detector-overpass observations (Section 3). We present the results as a progression: the first stage, the OLS and all-accident benchmarks, the clean free-flow-onset estimate, and the distance gradient from the road to surrounding households.

5.1 First stage

Table 4 reports the first stage for three instruments. Using all accidents, an accident in the prior two hours lowers the congestion index by 0.077 ($F \approx 1{,}293$); restricting to free-flow-onset accidents lowers it by 0.015 ($F \approx 48$); and using congested-onset accidents lowers it by 0.143 ($F \approx 1{,}849$). All three are strong by the Olea and Pflueger (2013) standard, but the event study (Section 4.3) shows that only the free-flow-onset coefficient reflects an exogenous shock—the congested-onset and all-accident coefficients are inflated by reverse causality. The free-flow-onset first stage is smaller because these accidents strike uncongested segments, but it clears the weak-instrument benchmark comfortably.

Table 4: First stage: freeway accidents and the congestion index

	All accidents	Free-flow onset	Congested onset
Accident, any onset (prior 2h)	-0.0768***		
	(0.00214)		
Free-flow-onset accident (prior 2h)		-0.0152***	
		(0.00218)	
Congested-onset accident (prior 2h)			-0.143***
			(0.00333)
First-stage F (cluster-robust)	1293.4	48.3	1849.4
Treated (instrument=1)	10,483	5,378	5,075
Observations (treated + clean controls)	3890495	3885390	3885087

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Dependent variable is the congestion index (speed / free-flow speed; lower = more congested). Detector (α_i), date (μ_d), and hour-of-day (λ_h) fixed effects; SE clustered at the detector. "All" uses every accident; "free-flow onset" restricts to accidents that struck while the segment was free-flowing (pre-accident CI ≥ 0.90), which the event study shows are not preceded by a congestion pre-trend; "congested onset" restricts to accidents during existing congestion, shown to document the reverse causality.

5.2 OLS and the contaminated instrument

Table 5 reports second-stage estimates. Within detector, date, and hour-of-day, the OLS coefficient on the congestion index is +0.14 ($p < 0.05$): small and wrong-signed, implying that more congestion is associated with very slightly *lower* temperature. (The sign and magnitude are unchanged when the local weather controls are added, so the OLS-to-IV sign flip below is not an artifact of omitting weather.) We read this not as evidence

against the mechanism but as a sign that cross-sectional confounding dominates the raw within-cell correlation—motivating the instrument. The all-accident 2SLS estimate is -1.14 ($p < 0.05$): correctly signed and precise. We do not, however, treat it as our preferred estimate, because the event study shows the all-accident instrument is contaminated by reverse causality. Adding contemporaneous traffic volume—a potential bad control—barely moves it (-1.19), so the all-accident result is not sensitive to that modeling choice; the binding problem is the instrument's validity, not volume.

Table 5: Effect of congestion on land surface temperature: OLS and 2SLS

	(1) OLS	(2) 2SLS (all)	(3) 2SLS (free-flow)	(4) 2SLS (+vol)
Congestion index	0.136**	-1.137**	-4.774	-1.187**
	(0.0590)	(0.493)	(3.477)	(0.499)
Traffic flow (robustness)				-0.0000699***
				(0.0000123)
Treated (instrument=1)	10,483	10,483	5,378	10,483
Observations (treated + clean controls)	3890495	3890495	3885390	3890495

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Outcome is LST (°C) in the 500-meter buffer; $\alpha_i + \mu_d + \lambda_h$ fixed effects; SE clustered at the detector. Columns: OLS; 2SLS with all accidents (contaminated by reverse causality, Section 4.3); 2SLS with free-flow-onset accidents (clean); and 2SLS adding contemporaneous volume (a potential bad control, Section 4.2.3). The free-flow-onset column is the binary instrument; Tables 6 and 7 refine it.

5.3 The clean free-flow-onset estimate

Restricting to the free-flow-onset instrument removes the reverse causality but costs precision, because these accidents move congestion less. The clean-IV sample comprises the free-flow-onset-treated overpasses together with clean no-accident control overpasses; congested-onset overpasses are dropped so they do not contaminate the control group. The "free-flow-onset treated" and "observations" rows in the clean-instrument tables report these two counts. Table 6 reports the progression. The binary free-flow-onset instrument gives $\delta = -4.77$ (SE 3.48), correctly signed but imprecise ($p = 0.17$). Weighting the instrument by accident duration—longer accidents sustain larger congestion—gives -6.02 ($p = 0.08$); using both instruments together (overidentified) gives -5.85 ($p = 0.06$). The Hansen J test does not reject ($p = 0.92$), which is consistent with—but only weak evidence for—instrument validity, since the overidentification test has limited power when the first stage is as modest as it is here. We therefore treat it as an internal consistency check rather than a test of the exclusion restriction.

Table 6: Clean free-flow IV: instrument variants + noise-cut outcome

	(1) ff binary	(2) ff duration	(3) ff overid	(4) demean
Congestion index	-4.774	-6.021*	-5.853*	-1.351*
	(3.477)	(3.388)	(3.062)	(0.780)
KP first-stage F	48.3	32.8	28.8	48.0
Hansen J p (overid)			0.921	
Free-flow-onset treated	5,378	5,217	5,217	5,334
Observations (treated + clean controls)	3885390	3885229	3885229	3866916

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Concentrating the outcome on the near-road area improves precision without attenuating the estimate. The odd-numbered columns of Table 7 report the overidentified clean instrument applied to tighter LST outcomes, before adding weather controls: at the 100-meter band the estimate is $\delta = -6.58$ (SE 3.01, $p = 0.029$); at 250 meters -6.15 ($p = 0.04$);

at 500 meters -5.85 ($p = 0.06$). The point estimate is stable across these outcomes, so the gain in significance comes from using more of the clean instrument information and a less noisy outcome, not from a change in the estimated effect.

Our preferred specification adds the local hourly weather controls $\mathbf{X}_{it}$—air temperature, relative humidity, wind speed, and precipitation at the detector and overpass hour. Table 7 reports the clean overidentified estimate with and without these controls across the near-road bands. The controls absorb a large share of residual temperature variance—the air-temperature coefficient alone is about +0.77°C of LST per °C of air temperature, so the congestion effect is identified *over and above* ambient temperature—and they attenuate the congestion estimate modestly while tightening its precision. With weather controls the estimate is $\delta = -5.36$ at 100 meters (SE 2.52, $p = 0.033$, first-stage $F \approx 30$, Hansen J $p = 0.83$), -4.97 at 250 meters ($p = 0.053$), and -4.30 at 500 meters ($p = 0.10$). The estimate stays right-signed and stable in rough magnitude, but significance at the conventional threshold survives only at the 100-meter band, with the 250-meter band at the margin. We take the weather-controlled 100-meter overidentified specification as the preferred two-stage least squares specification.

Table 7: Weather-controlled spec grid: overid (ff + duration) on near-road bands

	No weather (1) 100m	+ Weather (2) 100m	No weather (3) 250m	+ Weather (4) 250m	No weather (5) 500m	+ Weather (6) 500m
Congestion index	-6.581**	-5.357**	-6.148**	-4.974*	-5.853*	-4.298*
	(3.015)	(2.518)	(3.032)	(2.571)	(3.062)	(2.586)
KP first-stage F	29.5	29.7	29.3	29.3	28.8	28.9
Hansen J p (overid)	0.709	0.832	0.876	0.438	0.921	0.386
Free-flow-onset treated	4,954	4,699	5,108	4,843	5,217	4,944
Observations	3690632	3512527	3806820	3623207	3885229	3697660

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The quantity we emphasize, however, is the reduced form that underlies it: a free-flow-onset accident raises near-road surface temperature by about 0.07°C over the following

two hours (Section 4.3), against flat pre-accident leads. This is the direct experimental contrast—treated versus untreated overpasses of the same detector, date, and hour—and it requires no stand on how the instrument's association with congestion divides between the accident shock and the selection inherent in conditioning on a free-flowing pre-hour. The 2SLS coefficient converts this response into per-unit-congestion terms, and that conversion is where the remaining uncertainty lives: using the measured first stage (-0.015), the implied effect is roughly 0.7°C per 10 mph of speed reduction; under the selection-adjusted accounting from the placebo exercise in Section 6—which nets the pseudo-event response from both the first stage (giving ≈ -0.036) and the reduced form—it is roughly 0.3 to 0.4°C per 10 mph, where the small difference reflects whether the pseudo-event temperature response, itself statistically indistinguishable from zero, is netted at its point estimate or at zero. We therefore present the magnitude as a range, 0.3–0.7°C per 10 mph, which brackets the associational estimate of Lee and Berkelhammer (2025) (0.36°C per 10 mph). The corresponding free-flow-to-standstill figure—about 2.3 to 5.4°C (4.2 to 9.7°F)—is additionally a linear extrapolation across the full congestion range, which we interpret with care below (Section 5.4).

We report the full set of specifications rather than a single number. The clean estimate is right-signed throughout and statistically significant under the preferred near-road specification, but it is imprecise under the least demanding specification (binary instrument, 500-meter outcome, $p = 0.17$); readers should weight it accordingly. Because the design is overidentified and the first stage, though strong by the Olea and Pflueger (2013) benchmark, is not large ($F \approx 26$–30 on the weather-controlled specification), we also report Anderson–Rubin weak-instrument-robust confidence sets. For the preferred 100-meter estimate the 95% AR set is approximately $[-12.3, +0.5]$: it is wider than the Wald interval and marginally includes zero (the AR test does not reject a null effect at the 5% level, $p \approx 0.09$, though it does at 10%). Conventional inference thus places the near-road effect at the 5% level while fully weak-instrument-robust inference places it at the 10% margin. Because the AR

statistic is a test of the reduced-form relationship itself, we treat the weak-instrument-robust result as the operative statement of significance for the paper's central claim: the near-road temperature response is significant at the 10 percent level, not the 5 percent level.

5.4 Interpreting the magnitude

Two features of the design bound how the magnitude should be read. First, the free-flow-onset instrument identifies the effect from accidents on free-flowing segments, where the measured first-stage change in congestion is small (about -0.015 in the congestion index), and the placebo exercise of Robustness 4 shows that the causal first stage—measured against the no-accident path of a segment selected the same way—is larger, about -0.036. Because the 2SLS coefficient is the ratio of the reduced-form temperature response to the first stage, this accounting choice moves the per-unit magnitude by a factor of about two: roughly 0.7°C per 10 mph under the measured first stage, roughly 0.3 to 0.4°C per 10 mph under the causal one. The adjustment is applied consistently to both sides of the ratio: netting the pseudo-event response from the first stage alone gives -2.3 per unit of the congestion index, while netting it from the reduced form as well—the pseudo-event temperature response is -0.03°C, statistically indistinguishable from zero—gives -2.8, so the lower bound is insensitive to the treatment of the (null) pseudo reduced form. The reduced-form response itself is invariant to the choice, which is why we anchor interpretation there and report the per-unit magnitude as a range. Expressing either figure as a free-flow-to-standstill difference additionally extrapolates a locally identified response across the full congestion range; we treat the 2.3 to 5.4°C standstill figure as an illustrative linear extrapolation rather than a direct estimate. If the congestion-heat relationship is convex, with larger marginal effects under heavy congestion than near free-flow, the standstill figure overstates the average structural effect. The onset-threshold sensitivity analysis (Robustness 4) points the same way: the estimate is right-signed at every threshold defining "free-flow onset," but its magnitude moves with that definition while the first stage weakens, so we treat the point magnitude as

informative rather than sharply determined.

Second, the estimate is a local average treatment effect for the complier population—segments that are normally free-flowing and experience a sudden, exogenous slowdown—and need not equal the effect of chronic congestion on heavily trafficked corridors. This cuts in a specific direction: our compliers are precisely the segments where congestion is rare, so the estimate speaks to the marginal heat cost of an unusual slowdown rather than to the steady-state heat of a perpetually congested freeway. The large gap between the OLS association (+0.14) and the IV estimate partly reflects this small first stage, which mechanically magnifies the ratio, and partly reflects confounding: the wrong-signed OLS coefficient indicates that cross-sectional factors mask the effect in the raw correlation. A reader might reasonably worry that an instrument producing a sign flip of this size loads on a violation correlated with congestion rather than on the structural parameter. The event study is the relevant safeguard: congestion and surface-temperature pre-trends are both flat before free-flow-onset accidents (Section 4.3), which is the pattern a confounded instrument would not display.

Is the near-road effect just stopped vehicles in the pixel? Because ECOSTRESS records the radiometric temperature of whatever fills a pixel, a queue of sun-heated vehicles could in principle raise the measured near-road temperature without warming the surrounding environment—a compositional artifact rather than a thermal effect, and one that would be most acute in the 100-meter band where vehicles are densest. A back-of-envelope bounds this channel, and the key is that free-flow-onset accidents barely slow traffic. Under the measured first stage, the congestion-index change is about 0.015, which by the standard speed-density relation ($k/k_{\text{jam}} = 1 - \text{CI}$) corresponds to a density increase of about $0.015\,k_{\text{jam}}$. At a jam density of roughly 140 vehicles per kilometer per lane over five lanes, that is about ten extra vehicles per kilometer, or two within the ∼200-meter freeway segment that crosses the 100-meter buffer. At a projected area of about 10 m^2 each, these add ∼20 m^2 of vehicle

surface to the buffer's ∼31,000 m^2—a coverage increase of under 0.1 percent. Even if those surfaces are a generous 20°C hotter than their surroundings, the implied rise in the pixel-mean temperature is about 0.01°C, against a measured reduced-form response of roughly 0.07°C. The same arithmetic under the selection-adjusted accounting (first stage −0.036, with the correspondingly larger reduced form) yields about five vehicles and a bound near a third of the response. Vehicle composition can thus account for roughly a fifth to a third of the near-road effect under either accounting; the remainder reflects genuine warming of the road environment, consistent with the waste-heat mechanism. The same logic explains why the downwind and upwind half-disks give similar estimates (Robustness 5, Appendix): at these densities and buffer sizes neither wind-borne dispersion nor a compositional artifact is large enough to produce a detectable asymmetry.

5.5 From the road to nearby households

Table 8 and Figure 5 report the clean estimate by distance from the freeway. We use *disjoint* annular rings—0–100 meters, 100–500 meters, and 500–1,000 meters—rather than nested disks, so that each estimate reflects the temperature response in a distinct distance zone and the gradient is not mechanically smoothed by sharing near-road pixels across bands. The near-road ring captures on-road and roadside heat; the outer ring captures the residential area around the corridor, the exposure of most direct welfare interest. The gradient adjudicates two questions: whether a measurable effect persists out to the residential ring, and whether it decays with distance (consistent with heat spreading outward from the freeway) or instead rises in the intermediate zone (which would point to traffic diverted onto surface streets).

With weather controls, the gradient decays monotonically from the road outward (Table 8): $\delta = -5.36$ at the 0–100-meter ring ($p = 0.033$), -4.38 at 100–500 meters ($p = 0.08$), and -2.76 at the 500–1,000-meter residential ring ($p = 0.28$). The effect is right-signed in every ring but statistically resolved only near the road; the residential-ring estimate, though of the expected sign and a plausible magnitude, cannot reject zero. We therefore read

the near-road effect as robust and the temperature response in nearby residential areas as suggestive rather than established. The smooth decay—rather than a rise in the intermediate zone—is more consistent with heat spreading outward from the freeway than with a localized surface-street diversion effect, though the imprecision of the outer rings means we do not take a strong stand. The monotonic decay also serves as a within-corridor spatial falsification: an instrument loading on a broad, area-level confounder—regional weather or temperature anomalies correlated with accident timing—would raise temperature similarly at all distances rather than produce a clean decay toward the road.

Table 8: Weather-controlled distance gradient: disjoint rings (road to households)

	(1) 0–100m	(2) 100–500m	(3) 500–1000m
Congestion index	-5.357**	-4.383*	-2.757
	(2.518)	(2.481)	(2.566)
KP first-stage F	29.7	28.5	27.5
Hansen J p (overid)	0.832	0.688	0.309
Free-flow-onset treated	4,699	4,654	4,881
Observations	3512527	3459533	3648131

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

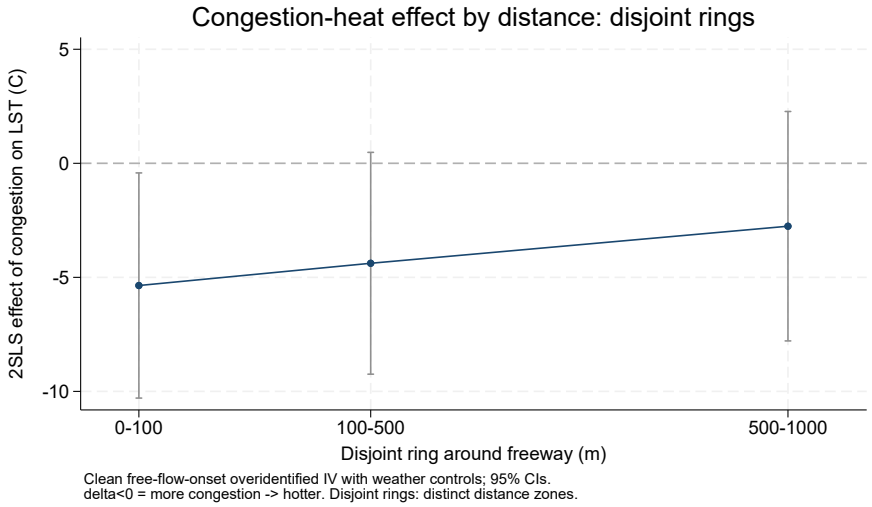

Figure 5: Congestion-heat effect by distance from the freeway. Clean free-flow-onset overidentified 2SLS estimate of δ (effect of the congestion index on LST, °C) with weather controls—the preferred specification—for disjoint annular rings 0–100, 100–500, and 500–1,000 meters from the detector, with 95% confidence intervals. Disjoint rings (rather than nested disks) ensure each estimate reflects a distinct distance zone. The outer ring covers the residential area around the corridor. Point estimates match Table 8.

How many people live in these rings? Table 9 reports the residential population around the studied freeway segments from the 2018–2022 American Community Survey. About 11 million Californians live within one kilometer of these segments, and 4.7 million within 500 meters. This population is more racially diverse than the state as a whole—42 percent Hispanic and 71 percent non-white, against roughly 40 and 65 percent statewide—though its median household income is close to the state median. This characterizes the scale and composition of the exposed population; it is not a claim that the congestion-heat effect falls disproportionately on these residents, which our estimates do not support.

Table 9: Residential population exposed near the studied freeway segments

Distance ring	Exposed pop.	Median income	% Hispanic	% non-white
0–100 m	319,927	$100,193	42	71
100–500 m	4,413,002	$100,329	42	71
500–1000 m	6,273,904	$102,833	41	70
Corridor (0–1000 m)	11,006,834	$101,752		
California	39,356,104	$104,733		

Notes: Population, median household income, and racial composition from the 2018–2022 ACS five-year block-group estimates, area-weighted into disjoint distance-to-nearest-freeway rings around the detectors in the estimation sample. Detector buffers are unioned across segments, so each resident is counted once, by distance to the nearest studied segment. "% non-white" is one minus the non-Hispanic white share; income is population-weighted across block groups. The California row is the statewide population-weighted benchmark.

6 Robustness Checks

This section reports the four checks most central to identification and the exclusion restriction: a prior-day placebo, a weather-orthogonality test, severity heterogeneity, and the onset-threshold sensitivity and pseudo-event placebo that probe the instrument's construction. Some checks are already reported with the main results—the distance-band gradient (Section 5, Figure 5) and the duration-weighted and overidentified instruments (Table 6). Seven further checks—two mechanism probes (wind dispersion, lookback window) and five specification checks (adjacent-segment exclusion, spatial demeaning, detector data quality, alternative LST aggregation, and the traffic-volume control), numbered 5 through 11—are summarized briefly at the end of this section, with the supporting tables collected in Appendix; none alters the conclusions.

1. *Placebo: Incidents from 24 hours prior.* We replace Z_{it} with an indicator for

whether a qualifying incident occurred on segment i at the same time of day on the previous calendar day. If the instrument is valid, prior-day incidents should have no predictive power for today's LST conditional on today's congestion. Table 10 confirms this: a prior-day free-flow-onset accident does not predict today's surface temperature (reduced-form coefficient -0.08°C, statistically insignificant). It carries a small, economically negligible same-sign correlation with today's congestion (day-to-day persistence in traffic conditions) but no temperature effect, as a valid instrument requires.

Table 10: Robustness: prior-day placebo

	(1) Congestion index	(2) LST (°C)
Prior-day ff accident	0.0085***	-0.0778
	(0.0016)	(0.0516)
Prior-day treated	5,329	5,329
Observations (no-accident sample)	3880012	3880012

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2. Weather orthogonality test. We regress the free-flow-onset instrument Z_{it} on the weather controls $\mathbf{X}_{it}$, holding the detector, date, and hour-of-day fixed effects fixed, to check that the instrument's timing is not systematically driven by weather that independently affects LST (Table 11). Weather explains essentially none of the within-detector, within-day variation in the instrument (within-$R^2 \approx 0.00$), and the coefficients are economically negligible: a one-degree increase in air temperature shifts the free-flow-onset rate by about 3×10^{-5} against a base rate near 1.3×10^{-3}. The four weather controls are jointly significant ($F = 2.73$, $p = 0.03$), but with 3.7 million observations even trivial correlations reject; the magnitudes, not the p-value, bear on the exclusion restriction, and they support it.

3. Severity heterogeneity. A natural severity weight is accident duration: longer accidents sustain larger congestion. The duration-weighted instrument, and the overidentified specification that combines it with the binary instrument, are reported in Table 6; the Hansen J test does not reject ($p = 0.92$), an internal consistency check that we read cautiously given

Table 11: Weather-orthogonality of the free-flow-onset instrument (Robustness)

	(1) ff_incident
Air temp (C)	0.00003**
	(0.00001)
Humidity (%)	0.00001**
	(0.00000)
Wind speed (m/s)	-0.00005
	(0.00003)
Precip (mm)	0.00002
	(0.00003)
Observations	3697817
Within R-sq	0.00002

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the low power of overidentification tests at this first-stage strength. A complementary split of the first stage by the number of lanes closed would provide further confirmation but awaits lane-closure detail in the incident records.

4. Free-flow-onset threshold sensitivity. The instrument conditions on a lagged value of the endogenous variable: an accident qualifies only if its segment was free-flowing (CI ≥ 0.90) in the hour before impact. Because this selects on transitory traffic states, the choice of threshold matters in two directions. Loosening it readmits accidents that congestion itself may have caused, reintroducing the reverse causality the restriction exists to remove; tightening it conditions on increasingly unusual hours, raising the worry that the selected moments are transitorily atypical in temperature as well, so that mean reversion in the outcome masquerades as a treatment response. Table 12 re-estimates the preferred specification at onset thresholds of 0.85, 0.90 (the baseline), and 0.95. The estimate is right-signed at every threshold, but its magnitude is sensitive to the choice: -2.1 and insignificant at 0.85, -5.4 at the baseline, and -11.7 ($p = 0.06$) at 0.95, with the first-stage F falling from 72 to 30 to 7 as the treated sample shrinks toward the most strictly free-flowing segments.

The loose end behaves as reverse-causality contamination predicts, attenuating toward the all-accident estimate of -1.14 (Table 5); the strict end is weakly identified ($F = 7.3$, far below conventional benchmarks), so its point estimate should not be read at face value. Against the mean-reversion reading, the event-study temperature leads for free-flow-onset accidents are flat individually and jointly ($p = 0.86$; Section 4.3): treated overpasses are not selected on transitorily warm or cool moments, so there is no temperature reversion for the post-accident bin to inherit.

We test the selection channel directly with a placebo that replicates the instrument's construction without an accident. For each real free-flow-onset accident we assign a pseudo-event at the same timestamp to a randomly drawn other detector that was free-flowing in the pre-hour (CI ≥ 0.90) and had no accident within six hours; the 283,236 pseudo-events reproduce the instrument's date and time-of-day distribution almost exactly. Table 13 reports their effects on the no-accident sample. The pseudo first stage is *positive* ($+0.021$, $t = 20$): a segment selected on a free-flowing pre-hour remains more free-flowing than its baseline over the following two hours—it persists; it does not revert below. The accident-induced drop to 0.015 *below* baseline therefore cannot be mean reversion, and measured against this no-accident counterfactual the accident's causal effect on congestion is about -0.036 rather than the -0.015 level difference. The pseudo reduced forms on temperature are null at both bands (-0.030 at 100 meters, $p = 0.53$; $+0.008$ at 500 meters, $p = 0.86$): the selection inherent in the instrument does not move surface temperature, which is direct support for the exclusion restriction. The placebo thus resolves the artifact question—the temperature response is caused by accidents, not by the conditioning—while making precise the first-stage accounting choice that motivates reporting the per-unit magnitude as a range (Section 5.4).

What the threshold gradient does establish is that the magnitude is not threshold-invariant. The sign and the qualitative conclusion are robust to the onset definition, but the point estimate is not, which reinforces our reading of the magnitude as informative at the

identified margin rather than sharply estimated (Section 5.4).

Table 12: Free-flow-onset threshold sensitivity (preferred specification: overidentified instruments, 100 m band, weather controls)

	(1) Onset CI ≥ 0.85	(2) ≥ 0.90 (baseline)	(3) ≥ 0.95
Congestion index	-2.089	-5.357**	-11.748*
	(1.654)	(2.518)	(6.267)
KP first-stage F	72.5	29.7	7.3
Free-flow-onset treated	5,853	4,699	2,990
Observations	3513681	3512527	3510818

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Placebo: pseudo free-flow-onset events (selection without an accident)

	(1) Congestion index	(2) LST 100m	(3) LST 500m
Pseudo free-flow-onset event	0.0208***	-0.0297	0.0082
	(0.0010)	(0.0476)	(0.0476)
Pseudo-event treated	5,693	5,406	5,693
Observations	3692716	3507828	3692716

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Additional checks. Seven further checks, with supporting tables in Appendix, leave the conclusions intact. (5) Splitting the 500-meter buffer into downwind and upwind half-disks by ERA5-Land wind direction yields no significant differential (-3.97 vs. -5.13), so we do not lean on a thermal-dispersion interpretation (Table 14). (6) Lookback windows of one and three hours, against the two-hour baseline, leave the estimate right-signed and comparable (-3.5 and -5.0), with the first-stage F falling as the window lengthens (Table 15). (7) Dropping observations on segments within two detector stations of an accident, to remove diverted-traffic contamination of the controls, leaves the estimate unchanged (-5.88; Table 16). (8) Spatially demeaning LST against nearby detectors sharply cuts the standard

error but attenuates the point estimate (Table 6, final column), indicating the effect is spatially broad rather than confined to a single buffer; we read it as such, not as a preferred specification. (9) Tightening the detector data-completeness threshold from 50 to 90 percent leaves the near-road estimate essentially unchanged (-5.7; Table 17). (10) Alternative LST aggregations—the nearest single pixel (-5.0) and the within-buffer maximum (-5.49, $p = 0.05$)— match the mean-based headline (Tables 18 and 19). (11) Adding the jointly determined traffic-volume control, a potential bad control, moves the congestion estimate negligibly (Table 5).

7 Conclusion

We estimate the short-run effect of freeway congestion on nearby surface temperature in California, using freeway accidents as an instrument for congestion and ECOSTRESS land surface temperature, matched to congestion at the same hour, as the outcome. The main lesson concerns identification. Accidents are not, in general, exogenous to congestion: an event study shows that congestion rises in the hours before a typical accident, so an instrument built from all accidents embeds reverse causality and yields a precise but contaminated estimate. Restricting the instrument to accidents that strike while the road is free-flowing removes this problem—congestion is flat before such accidents and jumps at impact—at the cost of a smaller first stage.

Under this clean design the estimated effect is correctly signed at every distance band. The central estimate is the reduced form: a free-flow-onset accident raises near-road surface temperature by about 0.07°C over the following two hours, against flat pre-trends, significant at the 10 percent level under weak-instrument-robust inference. Expressed per unit of congestion, the response implies roughly 0.3 to 0.7°C per 10 mph reduction in speed—a range whose bounds reflect how the instrument's first stage is apportioned between the accident shock and the selection inherent in its construction, and which brackets the

associational estimate of Lee and Berkelhammer (2025). We anchor the magnitude on this locally identified margin rather than on the larger free-flow-to-standstill extrapolation, which requires assuming the relationship is linear across the full congestion range, and we report the full specification grid rather than a single number. The effect decays with distance from the freeway: it is right-signed in the outer ring that contains nearby residential areas but is not statistically resolved there, so the temperature response experienced by nearby households is suggestive rather than established.

The estimates have clear limitations. The clean instrument is only borderline precise, so we view the magnitude as suggestive rather than tightly pinned down. Our preferred specification includes local hourly weather controls; they attenuate the estimate modestly and tighten its precision, leaving the near-road effect significant but the residential-ring response unresolved. We cannot fully separate heat from traffic diverted onto surface streets from heat spreading outward from the freeway; the distance gradient speaks to this, but we do not take a strong stand. The estimates are local to California freeway segments with PeMS detectors and to the variation induced by free-flow-onset accidents, and need not extend to other settings or to longer-run or larger congestion changes.

Within those bounds, the paper contributes a credibly identified estimate of a relationship that prior work could only document associationally, and a transparent way to handle reverse causality in accident-based instruments. Whether the near-road and household temperature responses documented here are large enough to matter for health or welfare, and how they compare to the cost of congestion relief, are questions we leave to future work.

References

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Barth, M. and Boriboonsomsin, K. (2008). Real-world carbon dioxide impacts of traffic congestion. *Transportation Research Record*, 2058(1):163–171.

Baum, C. F., Schaffer, M. E., and Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, 7(4):465–506.

Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2022). When is TSLS actually LATE? NBER Working Paper 29709, National Bureau of Economic Research.

California Department of Transportation (2026). Caltrans performance measurement system (PeMS). Station-hour traffic data and CHP incident logs, accessed for June 2018–March 2026.

Cao, D., Wu, J., Dong, X., Sun, H., Qu, X., and Yang, Z. (2021). Quantification of the impact of traffic incidents on speed reduction: A causal inference based approach. *Accident Analysis & Prevention*, 157:106163.

Chakraborty, T. et al. (2019). Disproportionately higher exposure to urban heat in lower-income neighborhoods: a multi-city perspective. *Environmental Research Letters*, 14(10):105003.

Chay, K. Y. and Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *Quarterly Journal of Economics*, 118(3):1121–1167.

Chay, K. Y. and Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy*, 113(2):376–424.

Currie, J. and Walker, R. (2011). Traffic congestion and infant health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1):65–90.

Declet-Barreto, J., Brazel, A. J., Martin, C. A., Chow, W. T., and Harlan, S. L. (2013). Creating the park cool island in an inner-city neighborhood: heat mitigation strategy for phoenix, AZ. *Urban Ecosystems*, 16(3):617–635.

Deschenês, O. and Greenstone, M. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1):354–385.

Duranton, G. and Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, 101(6):2616–2652.

Fisher, J. B. et al. (2020). ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. *Water Resources Research*, 56.

Gago, E. et al. (2013). Urban heat islands: mitigation strategies. *Renewable and Sustainable Energy Reviews*, 25:749–758.

Graff Zivin, J. and Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–3673.

Greenstone, M. and Gayer, T. (2018). Quasi-experimental and experimental approaches to environmental economics. Technical Report Working Paper 24903, National Bureau of Economic Research.

Haule, H. J., Sando, T., Lentz, R., Chuan, C.-H., and Alluri, P. (2019). Evaluating the impact and clearance duration of freeway incidents. *International Journal of Transportation Science and Technology*, 8(1):13–24.

He, F., Yan, X., Liu, Y., and Ma, L. (2016). A traffic congestion assessment method for urban road networks based on speed performance index. *Procedia Engineering*, 137:425–433.

Health Effects Institute (2010). Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects. Technical Report Special Report 17, Health Effects Institute, Boston, MA.

Heaviside, C., Macintyre, H., and Vardoulakis, S. (2017). The urban heat island: implications for health in a changing environment. *Current Environmental Health Reports*, 4(3):296–305.

Hulley, G. and Hook, S. (2022). ECOSTRESS tiled land surface temperature and emissivity instantaneous L2 global 70 m V002. Data set, accessed via AppEEARS.

Husni, E. et al. (2022). Microclimate investigation of vehicular traffic on the urban heat island through IoT-based device. *Heliyon*.

Ichinose, T., Shimodozono, K., and Hanaki, K. (1999). Impact of anthropogenic heat on urban climate in Tokyo. *Atmospheric Environment*, 33(24–25):3897–3909.

Knittel, C. R., Miller, D. L., and Sanders, N. J. (2016). Caution, drivers! Children present: Traffic, pollution, and infant health. *Review of Economics and Statistics*, 98(2):350–366.

Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. (2022). Valid t-ratio inference for IV. *American Economic Review*, 112(10):3260–3290.

Lee, J. and Berkelhammer, M. (2025). Evaluating the influence of traffic congestion on surface urban heat intensity. *Geophysical Research Letters*.

Li, Y. et al. (2020). Socioeconomic drivers of urban heat island effect in major Chinese cities. *Sustainable Cities and Society*.

LSE Grantham Research Institute (2022). Congestion pricing, air pollution, and individual-level behavioural responses. Technical Report Working Paper 362, London School of Economics.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725–753.

NASA Land Processes Distributed Active Archive Center (2024). ECOSTRESS ECO_L2T_LSTE version 2: product documentation and mission status. NASA LP DAAC. Mass storage unit anomalies suspended data acquisition 2018-09-29 to 2018-12-05 and 2019-03-14 to 2019-05-15.

National Weather Service (2024). Weather related fatality and injury statistics. National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Accessed June 2026.

Oke, T. (1973). City size and the urban heat island. *Atmospheric Environment*, 7(8):769–779.

Olea, J. L. M. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Quah, A. K. L. and Roth, M. (2012). Diurnal and weekly variation of anthropogenic heat emissions in a tropical city, Singapore. *Atmospheric Environment*, 46:92–103.

Sailor, D. J. (2011). A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. *International Journal of Climatology*, 31(2):189–199.

Sailor, D. J. and Lu, L. (2004). A top–down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. *Atmospheric Environment*, 38(17):2737–2748.

Santamouris, M., Cartalis, C., Synnefa, A., and Kolokotsa, D. (2015). On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. *Energy and Buildings*, 98:119–124.

Tran, M.-T., Swindell, C., Yang, B., Skupin, A., and Akerkar, R. (2026). Quantifying the impact of transportation on urban heat islands. *Annals of the American Association of Geographers*, 116(5):1190–1210.

Wilson, B. (2020). Urban heat management and the legacy of redlining. *Journal of the American Planning Association*, 86(4):443–457.

Xia, F., Cheng, X., Lei, Z., Xu, J., Liu, Y., Zhang, Y., and Zhang, Q. (2023). Heterogeneous impacts of local traffic congestion on local air pollution within a city: Utilizing taxi trajectory data. *Journal of Environmental Economics and Management*, 122:102896.

Yang, B. et al. (2026). Quantifying highway expansion impact on urban heat island effect in San Francisco Bay Area. *Cities*.

Appendix Additional Figures and Tables

Figures 6 and 7 show the all-accident event study discussed in Section 4.3: congestion rises in the hours before a typical accident (reverse causality), while surface-temperature leads are flat. The remaining inputs are the supporting tables for robustness checks 5 through 11, summarized in Section 6; none alters the conclusions of the main text. (Checks 8 and 11 are reported with main-text tables—the spatial-demeaning estimate in Table 6 and the volume-augmented specification in Table 5—and so have no separate table here.)

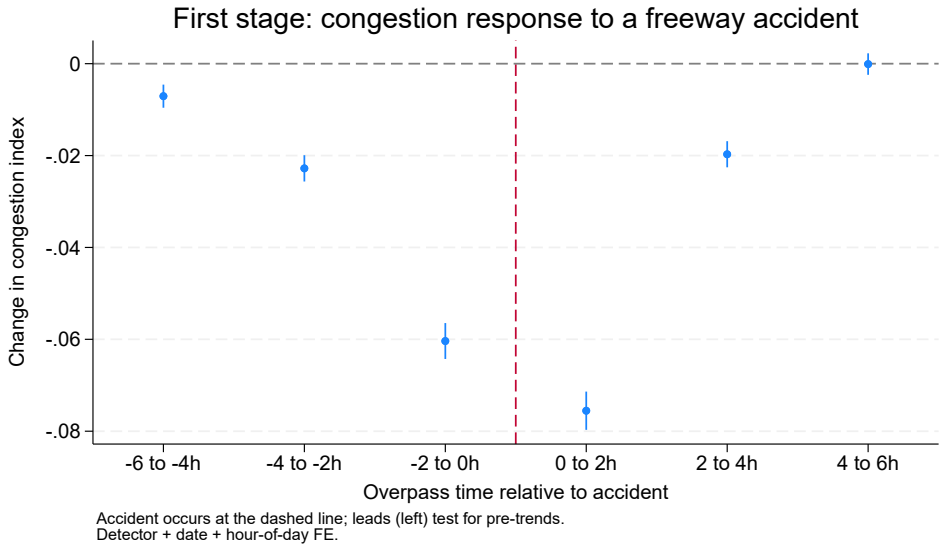

Figure 6: Event study, all accidents: congestion index by time relative to the accident. Estimated $\hat{\phi}_{\ell,u}$ from equation (1), Y = CI. Detector + date + hour-of-day fixed effects; 95% CIs clustered at the detector; baseline = no accident within ±6 hours. Congestion rises before the accident (negative leads), indicating reverse causality.

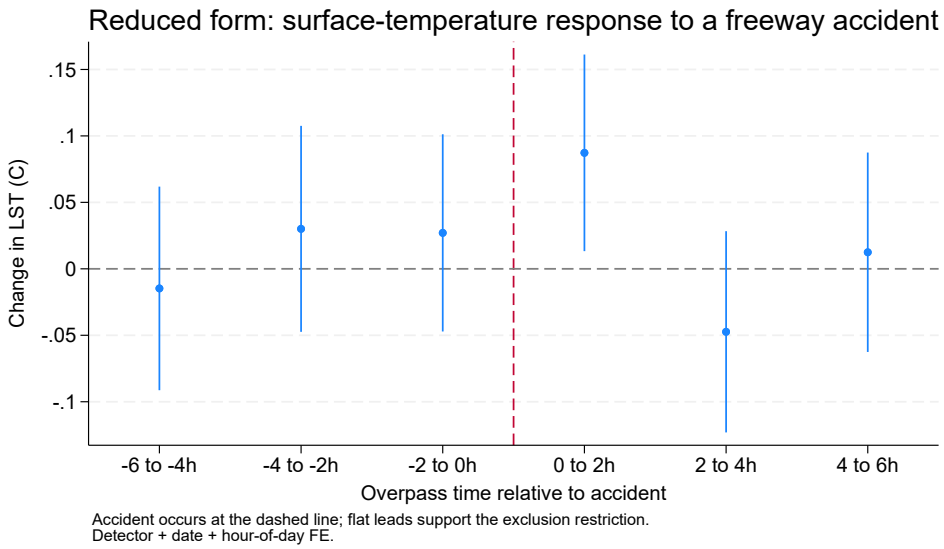

Figure 7: Event study, all accidents: surface temperature by time relative to the accident. Same specification as Figure 6, $Y = \text{LST}$ (°C). Pre-accident leads are flat and insignificant; the only significant response is at the accident bin.

Table 14: Robustness: downwind vs upwind 500m half-disk (weather-controlled)

	(1) Downwind	(2) Upwind
Congestion index	-3.974	-5.129**
	(2.563)	(2.606)
KP first-stage F	29.1	28.9
Free-flow-onset treated	4,908	4,906
Observations	3667658	3667191

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Robustness: lookback window sensitivity (100-meter outcome)

	(1) h = 1 hour	(2) h = 3 hours
Congestion index	-3.545	-5.011
	(3.494)	(4.088)
KP first-stage F	44.2	36.0
Free-flow-onset treated	2,594	7,658
Observations (treated + clean controls)	3688272	3690801

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Robustness: adjacent-segment exclusion (headline 100m, weather-controlled)

	(1) Full sample	(2) Drop +/-2 neighbors
Congestion index	-5.357**	-5.876**
	(2.518)	(2.658)
KP first-stage F	29.7	26.7
Free-flow-onset treated	4,699	4,558
Observations	3512527	3500300

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Robustness: detector data completeness tightened to 90 percent

	(1) 100m (90% completeness)	(2) 500m (90% completeness)
Congestion index	-5.661*	-4.541
	(3.122)	(3.249)
KP first-stage F	56.8	55.0
Free-flow-onset treated	4,257	4,489
Observations (treated + clean controls)	3136191	3302563

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Robustness: nearest-pixel LST aggregation

	(1) Nearest pixel
Congestion index	-4.970
	(3.535)
KP first-stage F	43.9
Free-flow-onset treated	4,828
Observations (treated + clean controls)	3470881

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 19: Robustness: maximum LST in the 500m buffer (weather-controlled)

	(1) Max LST (500m)
Congestion index	-5.487**
	(2.796)
KP first-stage F	28.9
Hansen J p	0.465
Free-flow-onset treated	4,944
Observations	3697641

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$