A Data construction and reproducibility

A.1 Sample assembly

The source panel begins with the public 2018 province-hour load data. Hourly records are aggregated to province-day and matched one-to-one to daily coal-power emissions and source weather by province and date. Static pre-period variables are then matched many-to-one by province. The merge retains 10,950 observations, exactly 30 provinces times 365 days. No load, emissions, wind, or weather values are imputed.

The receptor panel begins with all ordered source-receptor-day pairs. Pair geometry supplies distance and source-to-receptor bearing. Daily source wind determines whether a receptor falls inside each directional construction. Contemporaneous source emissions form the endogenous transport indices; source-specific mean 2017 emissions form the instruments. Pair contributions are summed to receptor-day and matched one-to-one to population-weighted CHAP $PM_{2.5}$ and receptor weather. The 2018–2020 panel contains 32,880 observations, exactly 30 receptors times 1,096 days.

Table 5: Variable dictionary for submitted analysis files

Variable	Definition	Empirical role
`cooling_shift_c24_ac2016`	CDD24 multiplied by 2016 AC ownership in appliances per household	Excluded source-day cooling-demand shifter
`log_daily_load_mwh`	Natural log of observed daily provincial electricity load	Endogenous source-day load
`pm25_mg`	Daily primary PM2.5 emissions from coal-fired power plants, Mg	Preferred physical source outcome
`wind_iv_60`	Sum of fixed 2017 source PM2.5 scale aligned within 60 degrees	Excluded receptor-day transport shifter
`transport_60_pm25_mg`	Sum of contemporaneous source PM2.5 aligned within 60 degrees	Endogenous receptor-day transport
`log_chap_pm25_popw`	Log population-weighted receptor PM2.5	Receptor outcome
`marginal_damage_2020_usd_per_gwh`	Downwind damage from one marginal GWh on hot source-days	Preferred welfare-accounting estimand

A.2 Vulnerability index

The vulnerability index is not part of either identification strategy. It is used only to describe receptor incidence. Each component is standardized across the 30 provinces and oriented so that a larger value denotes greater vulnerability. The index is the unweighted mean of the five z-scores. The high-vulnerability group is the upper tercile.

Table 6: Construction of the provincial vulnerability index

Component	Construction	Orientation	Weight
Income	Pre-period disposable income per capita	Lower is worse	0.20
Cooling access	Pre-period air conditioners per 100 households	Lower is worse	0.20
Old-age dependency	Population above working age relative to working-age population	Higher is worse	0.20
Urban-rural cooling gap	Difference in AC ownership between urban and rural households	Higher is worse	0.20
Heat-load sensitivity	Province-specific response of observed 2018 load to cooling degree days	Higher is worse	0.20

Notes: Each component is converted to an oriented z-score before averaging. The index is used for descriptive welfare incidence only and does not enter either excluded instrument or estimating equation.

B Source-side checks

B.1 Threshold and baseline grid

The preferred cooling threshold and AC baseline were fixed before the final production run. Table 7 reports every threshold and baseline variant used in the design. Coefficients across thresholds are not directly comparable in the first stage because a unit of CDD changes with the threshold, but the corresponding IV estimates answer the same load-to- emissions question. The 2010 baseline uses urban-household ownership and therefore changes both timing and household scope.

Table 7: Source-side threshold and baseline sensitivity

Specification	Outcome	First stage	F	Reduced form	2SLS
threshold: 18	$PM_{2.5}$ (Mg)	0.019	13.6	1.011^{**}	53.380^{**}
threshold: 18	$\log(PM_{2.5})$	0.019	13.6	0.018^{*}	0.937^{*}
threshold: 20	$PM_{2.5}$ (Mg)	0.025	14.7	1.344^{**}	54.540^{**}
threshold: 20	$\log(PM_{2.5})$	0.025	14.7	0.023^{*}	0.950^{*}
threshold: 22	$PM_{2.5}$ (Mg)	0.033	15.0	1.837^{**}	56.470^{**}
threshold: 22	$\log(PM_{2.5})$	0.033	15.0	0.031^{**}	0.966^{**}
threshold: 24	$PM_{2.5}$ (Mg)	0.044	15.1	2.553^{**}	58.439^{**}
threshold: 24	$\log(PM_{2.5})$	0.044	15.1	0.042^{*}	0.970^{*}
AC baseline: 2010 (urban households)	$PM_{2.5}$ (Mg)	0.034	27.1	0.836	24.842
AC baseline: 2015	$PM_{2.5}$ (Mg)	0.045	14.3	2.167^{*}	47.896^{*}
AC baseline: 2016	$PM_{2.5}$ (Mg)	0.044	15.1	2.553^{**}	58.439^{**}

Notes: All specifications use the controls and fixed effects in Table 2. Standard errors and exact p-values are available in the machine-readable source file.

B.2 Load outcomes and timing

Table 8 reports the daily load outcomes and contrasts from the saturated hourly model. Each hourly equation contains the same controls and absorbed effects as the preferred daily first stage. The contrasts are linear combinations from one joint interaction model.

Table 8: Load outcomes and hourly contrasts

Outcome or contrast	Coefficient	Standard error	F
Daily total load	0.044^{***}	(0.011)	15.1
Daily peak load	0.051^{***}	(0.011)	20.1
Afternoon-evening load	0.047^{***}	(0.011)	17.8
Overnight load	0.039^{***}	(0.012)	10.9
20:00 minus 05:00	0.019^{***}	(0.005)	–
21:00 minus 04:00	0.017^{***}	(0.004)	–
Afternoon-evening minus overnight	0.008^{**}	(0.003)	–

Notes: Load outcomes are in logs. Hourly contrasts come from a single saturated hour-interaction model, rather than subtraction of separately estimated coefficients.

B.3 Falsification and leverage

Future cooling shifts enter jointly with future CDD levels and squares. This matters because tomorrow's heat is correlated with today's heat. The falsification asks whether the AC-weighted component of future heat predicts current emissions after generic future heat has been absorbed. The answer is no. The same future terms remain jointly predictive of current load, a result consistent with multi-day heat spells rather than reverse causality from future demand.

Table 9: Cooling-shift timing and heating falsifications

Outcome	Term	Coefficient	Standard error	p-value
$PM_{2.5}$ (Mg)	Current cooling shift	3.047^{***}	(0.909)	0.001
$PM_{2.5}$ (Mg)	Cooling shift at t+1	-0.066	(0.184)	0.720
$PM_{2.5}$ (Mg)	Cooling shift at t+3	-0.207	(0.214)	0.334
$PM_{2.5}$ (Mg)	Cooling shift at t+7	-0.373	(0.435)	0.391
log(daily load)	Current cooling shift	0.024^{***}	(0.006)	0.000
log(daily load)	Cooling shift at t+1	0.007^{***}	(0.003)	0.004
log(daily load)	Cooling shift at t+3	0.010^{**}	(0.004)	0.014
log(daily load)	Cooling shift at t+7	0.008^{*}	(0.004)	0.077
log(daily load)	HDD x AC	-0.003	(0.003)	0.249
$PM_{2.5}$ (Mg)	HDD x AC	0.186	(0.232)	0.421

Notes: Future cooling shifts enter jointly with corresponding future generic-heat controls. Their joint p-value in the current-emissions equation is 0.570. Future heat remains predictive of current load, consistent with persistent hot spells; this is why the emissions falsification, not the load falsification, is the exclusion diagnostic.

The leave-one-source analysis re-estimates all fixed effects and controls after each omission. Table 10 summarizes its complete range; the underlying 30 estimates are included in the submitted source data.

Table 10: Leave-one-source-out stability

Across 30 omissions	First-stage F	Reduced form	2SLS
Minimum	10.7	1.574	36.748
Median	14.7	2.551	58.420
Maximum	36.6	3.464	91.084

Notes: Each estimate omits one source province and re-estimates the full preferred source specification. Every reduced-form and 2SLS coefficient is positive.

C Transport checks

C.1 Alternative wind, cones, and distance

Table 11 changes one element of the transport construction at a time. Station and NCEP indices are standardized separately, so coefficients compare a one-standard-deviation shift under each wind product. The inverse-distance construction divides a downwind source's contribution by source-receptor distance, with distance bounded below at 100 km. The two distance-window specifications exclude sources farther than 500 or 1,000 km.

Table 11: Wind-IV construction robustness

Wind data	Construction	First-stage F	Reduced form	2SLS
Station	45-degree cone	3435.2	0.031***	0.033***
Station	60-degree cone	3207.9	0.034**	0.036**
Station	90-degree cone	3126.9	0.032**	0.036**
Station	Isolated switch	8452.9	0.008*	0.008*
Station	Inverse distance	511.1	0.036**	0.035**
Station	≤500 km	755.4	0.018**	0.018**
Station	≤1,000 km	4426.6	0.032***	0.034***
NCEP	45-degree cone	2964.1	0.019*	0.020*
NCEP	60-degree cone	3303.3	0.025**	0.027**
NCEP	90-degree cone	3277.2	0.036***	0.040***
NCEP	Isolated switch	12374.3	0.005	0.005

Notes: All outcomes are log population-weighted receptor $PM_{2.5}$. Distance constructions use the station wind measure.

C.2 Direction and event time

Direction falsifications preserve the preferred controls and absorbed effects. Upwind sectors use source winds pointing away from the receptor. Rotations recalculate alignment after adding 90, minus 90, or 180 degrees to the observed wind-to direction. The wind-switch outcome activates when a pair enters the preferred downwind cone; the isolated-switch outcome further requires that it exits on the following day.

Table 12: Directional falsifications

Construction	Reduced-form coefficient	Standard error	p-value
60	0.034**	(0.013)	0.012
upwind60	-0.021	(0.016)	0.188
upwind45	-0.021	(0.015)	0.163
crosswind	-0.008	(0.005)	0.107
rot90	0.023	(0.017)	0.174
rotm90	-0.022*	(0.012)	0.078
rot180	-0.021	(0.016)	0.188
switch	0.001	(0.006)	0.808
isolated	0.008*	(0.004)	0.078

Notes: Each row replaces the preferred direction construction while preserving its controls and absorbed effects.

Event-time coefficients are estimated on a common 32,460-observation sample so changes across horizons do not reflect missing endpoint days. The spaced pre-event test was specified as the primary timing gate because its horizons are less collinear. For completeness, the denser test adds day -2 and rejects the joint null even though each included lead is individually insignificant.

Table 13: Wind-IV event time

Horizon	Coefficient	Standard error	95% interval
-7	0.005	(0.011)	[-0.016, 0.027]
-3	0.008	(0.011)	[-0.014, 0.029]
-2	0.003	(0.010)	[-0.017, 0.024]
-1	0.009	(0.010)	[-0.009, 0.028]
+0	0.034^{**}	(0.013)	[0.008, 0.060]
+1	0.041^{**}	(0.019)	[0.002, 0.079]
+2	0.020	(0.012)	[-0.004, 0.044]
+3	0.010	(0.008)	[-0.007, 0.026]
+7	0.008	(0.011)	[-0.013, 0.029]

Notes: The spaced pre-event horizons -7, -3, and -1 have joint $p = 0.710$. Adding horizon -2 yields joint $p = 0.018$, despite each individual lead being insignificant. Event day and the first lag have joint $p = 0.042$.

Table 14: Leave-one-receptor-out stability

Across 30 omissions	First-stage F	2SLS coefficient	p-value
Minimum	2735.9	0.029	0.001
Median	3098.8	0.036	0.013
Maximum	4161.7	0.045	0.042

Notes: Each estimate omits one receptor and re-estimates the preferred wind-IV specification.

D Welfare construction

D.1 Damage-factor transfer

For each year from 2017 through 2020, national particulate-emission damage in current US dollars is converted to 2020 dollars with the World Bank consumer price index. Dividing by annual mean population-weighted $PM_{2.5}$ yields damage per one $\mu g/m^3$-year. The preferred factor is the mean across the four years; the minimum and maximum enter the conservative sensitivity range. Receptor allocation uses the same factor for every province, so cross-province differences arise from predicted concentration changes and population rather than assumed differences in the value of health.

D.2 Uncertainty

The parametric simulation draws the source IV coefficient and wind IV coefficient independently from normal distributions centered at their point estimates with their clustered standard errors. It samples the four observed annual damage factors with replacement. Each of 20,000 draws rescales the preferred marginal damage. This procedure is transparent but uses normal approximations and does not model covariance between estimates taken from different samples.

The weak-IV/damage-factor envelope takes the Cartesian product of the source and wind Anderson–Rubin endpoints and the minimum and maximum damage factors. It is deliberately conservative and is not assigned a nominal joint coverage probability.

D.3 Annual proportional sensitivity

The main text reports only the marginal external cost because it remains on observed hot-day load support. An annual sensitivity uses the log-emissions reduced form to define a cooling-related fraction of observed source emissions:

$$E_{st}^{cool} = E_{st}^{obs}\left[1 - \exp\left(-\widehat{\rho}_{\log E} CDD24_{st} AC_{s,2016}\right)\right]. \tag{13}$$

This construction cannot assign more cooling-related emissions than the emissions observed on a source-day. Applying the same wind routing and damage factor gives \$100.99 million in 2020 dollars for 2018. The number is an annual proportional-emissions sensitivity, not the paper's primary estimand.

An earlier physical-level integration applied the homogeneous Mg coefficient to the entire cooling-shift distribution. In low-coal provinces, that calculation could predict a cooling component larger than observed emissions. It failed the support check and is not reported as a welfare result.

D.4 Complete receptor allocation

Table 15 reports every receptor contribution. Values sum to the national marginal estimate apart from rounding.

Table 15: Complete receptor-level marginal welfare incidence

Rank	Receptor	USD/GWh	Damage share (%)	Population share (%)	Ratio
1	Henan	188.1	9.11	6.89	1.322
2	Shandong	174.6	8.46	7.21	1.173
3	Hebei	174.6	8.45	5.42	1.560
4	Sichuan	137.0	6.63	5.98	1.109
5	Jiangsu	115.7	5.60	5.79	0.968
6	Anhui	93.9	4.55	4.51	1.009
7	Guangdong	93.4	4.52	8.05	0.562
8	Hubei	87.1	4.22	4.25	0.992
9	Shanxi	85.7	4.15	2.67	1.556
10	Hunan	83.7	4.05	4.94	0.820
11	Shaanxi	71.0	3.44	2.76	1.244
12	Xinjiang	68.6	3.32	1.76	1.886
13	Liaoning	61.0	2.95	3.15	0.938
14	Beijing	60.2	2.92	1.56	1.864
15	Yunnan	55.6	2.69	3.46	0.778
16	Guangxi	55.1	2.67	3.52	0.759
17	Zhejiang	50.0	2.42	4.08	0.594
18	Gansu	46.5	2.25	1.89	1.190
19	Guizhou	46.3	2.24	2.58	0.870
20	Inner Mongolia	45.0	2.18	1.82	1.197
21	Chongqing	43.2	2.09	2.22	0.945
22	Jiangxi	43.0	2.08	3.33	0.625
23	Heilongjiang	39.5	1.91	2.73	0.700
24	Tianjin	34.2	1.65	1.12	1.474
25	Jilin	30.1	1.46	1.96	0.744
26	Fujian	26.1	1.26	2.82	0.448
27	Shanghai	24.8	1.20	1.74	0.690
28	Qinghai	14.6	0.71	0.63	1.128
29	Ningxia	13.4	0.65	0.49	1.325
30	Hainan	3.0	0.14	0.67	0.216

Notes: Receptor values sum to the national cooling-relevance-weighted marginal estimate in Table 4.