Estimating the Economic Impacts of Climate Stress on U.S. Dairy Production: A Debiased Machine Learning Approach

Duong P. Ta
duongta@vt.edu

Mario Ortez
omario@vt.edu

Bradford Mills
bfmills@vt.edu

Michael Cary
macary@vt.edu

Department of Agricultural and Applied Economics, Virginia Tech

Acknowledgments: The authors gratefully acknowledge the Department of Agricultural and Applied Economics at Virginia Tech for supporting this project by providing research facilities.

Funding: This research was funded by The Dairy Alliance under the grant titled "Enhancing dairy sector resilience through data-driven infrastructure planning and forecast modeling."

Conflict of Interest: The authors declare that they have no competing interests.

Abstract

We estimate how climate stress, operating through cold and heat exposure, affects two key aspects of U.S. dairy production: per-cow milk volume and milk fat concentration. The analysis combines USDA NASS outcomes with hourly ERA5-Land weather in a monthly panel of 21 major dairy states from 2003 through 2025. State-level exposure is aggregated from county-level weather using milk-cow inventory weights. Cumulative current-month through four-month-lagged average partial derivatives are estimated using panel Automatic Debiased Machine Learning (ADML). The preferred neural-network ADML estimates imply that one additional heat-stress day with Temperature Humidity Index (THI) at or above 69 reduces milk production per cow by 0.319 percent and fat content by 0.00246 percentage points over the response window. Cold stress below THI 39 also reduces milk volume, by 0.0782 percent, but raises fat content by 0.00443 percentage points per exposure day. A semi-synthetic Monte Carlo calibrated to the empirical panel gives the preferred estimator the lowest root mean squared error, although 95 percent confidence intervals achieve approximately 50 percent coverage under 21 state clusters in simulation, requiring that uncertainty intervals be interpreted as indicative bounds. Applying the quantity effects to observed exposure, annual NASS production, and prices yields partial-equilibrium sensitivity estimates of $16.233 billion in heat-related and $4.480 billion in cold-related gross milk-revenue losses across the contiguous United States during 2003-2025. Year-specific effects suggest attenuation of the per-cow heat sensitivity coefficient over time, although expanding thermal exposure keeps aggregate industry damages elevated; specific adaptation mechanisms are not identified in the analysis.

1 Introduction

The U.S. dairy sector supports more than 3 million jobs and nearly $780 billion in annual economic activity (International Dairy Foods Association (IDFA), 2025). Dairy production also anchors rural demand for labor and services, while milk provides concentrated sources of calcium, protein, and vitamin D (A. Garcia and K. Kalscheur, 2004; H. Górska-Warsewicz et al., 2019; R. P. Heaney, 2014). This economic and nutritional role has been further amplified by growing consumer demand and commercial processing for protein-dense dairy products (e.g., USDA Economic Research Service (2023); OECD/FAO (2022)), heightening the economic stakes of climate-driven yield and component shocks. How national milk production and fat content respond to the thermal environment determines the stability of this sector.

Temperature stress alters both milk volume per cow and milk composition. Heat suppresses feed intake and induces metabolic and endocrine responses that reduce mammary productivity (C. T. Kadzere et al., 2002; J. W. West, 2003). Heat can also lower fat and protein measures that matter for manufacturing yield and component-based payment (W. Vroege et al., 2023; S. S. Moore et al., 2023). Cold creates a different energy allocation problem: cows must divert nutrients toward thermoregulation, which can lower milk volume even when seasonal feeding and concentration effects raise measured fat content. A complete economic assessment must therefore examine both tails of the Temperature-Humidity Index (THI) distribution and keep quantity and composition on their distinct scales.

Identifying climate stress effects is difficult even though short-run weather is not chosen by producers. Thermal exposure follows strong seasonal patterns and differs across states along with local herd geography, productivity, genetics, and cooling investments that have further changed over time. The response can also be nonlinear, interact with moisture, and persist beyond the month of exposure. A hand-selected linear fixed-effects regression may therefore confuse functional-form error with a causal response. We address this latter challenge using panel Automatic Debiased Machine Learning (ADML) with orthogonal Riesz scores and state-blocked cross-fitting to remove unobserved unit confounding and flexibly capture nonlinear response surfaces without restrictive parametric assumptions.

This paper makes five contributions. First, we apply panel ADML to livestock productivity and estimate a common two-tail design for milk yield and fat content over a 23-year

monthly panel. Second, we aggregate county weather using milk-cow inventories, aligning measured exposure with the disaggregated spatial location of production. Third, we document a sharp cross-margin asymmetry: heat reduces both volume and fat content, whereas cold reduces volume but raises fat concentration. Fourth, we evaluate the estimator in a semi-synthetic Monte Carlo that retains the actual weather panel, favoring NNET ADML for point estimation while clarifying finite-sample coverage limits. Fifth, we evaluate temporal heterogeneity across 2003–2025 and show that while the per-cow heat sensitivity coefficient has attenuated over time, expanding annual heat exposure and industry scale cause aggregate milk losses to rise, demonstrating that adaptation gains have not kept pace with compounding thermal exposure.

The preferred results imply a 0.319 percent cumulative decline in milk production per cow and a 0.00246 percentage-point decline in fat content for one additional `THI69_Days` day. The corresponding cold effects are a 0.0782 percent decline in production per cow and a 0.00443 percentage-point increase in fat content. We then translate the quantity effects, but not the composition effects, into a partial-equilibrium valuation using observed exposure, state production, and all-milk prices. The remaining sections review related evidence, describe the data and estimator, present validation and empirical results, evaluate adaptation and economic magnitude, and conclude with implications and limitations.

2 Related literature and contribution

2.1 Animal-science dose response

Animal-science research provides the physiological mechanism and threshold ranges that motivate our exposure design. Dairy cows generate substantial metabolic heat through rumen fermentation and lactation. When heat dissipation is constrained, respiration and body temperature rise, feed intake falls, and milk synthesis declines (D. V. Armstrong, 1994; C. T. Kadzere et al., 2002; J. W. West, 2003). Historical work often used THI 72 as the onset of heat stress, but evidence for modern high-producing cows places the threshold closer to 68 or 69 (T. M. McWhorter et al., 2023). We therefore define heat exposure at THI 69 rather than waiting for more severe conditions.

Cold stress receives less attention, partly because lactating cows generate substantial

internal heat. Yet the lower tail matters in large observational panels. U.S. evidence shows productivity losses below approximately THI 39 and a gradual cold-response gradient rather than a single sharp break (M. Gisbert-Queral et al., 2021; E. Choi et al., 2024). We use THI 39 as the lower boundary of the reference range. These thresholds summarize exposure rather than claim that every breed, housing system, or lactation stage shares an identical biological cutoff.

2.2 Measuring climate stress in agricultural economics

Agricultural climate-economics commonly represents nonlinear thermal exposure in two related ways. The first records time spent in temperature bins or beyond a threshold. W. Schlenker and M. J. Roberts (2009) reconstruct the within-day temperature distribution across narrow intervals, while E. Zaveri and D. B. Lobell (2019) measure extreme heat as the cumulative number of growing-season days above 30 degrees Celsius. J. Tack et al. (2015) similarly combine duration below a freezing threshold with heat degree-day variables. These distribution-based measures preserve damaging tails that seasonal averages can obscure (S. Hsiang, 2016).

The second approach weights exposure by its distance from a threshold. Crop studies call these measures growing, harmful, exceedance, or extreme degree days. W. Schlenker and M. J. Roberts (2009) integrate temperature degrees above crop-specific damage thresholds; F. Annan and W. Schlenker (2015) and M. Burke and K. Emerick (2016) use degree days above 29 degrees Celsius to study heat damages and adaptation in corn and soybeans; and J. Tack et al. (2015) use interpolated degree days to measure wheat exposure. Their repeated use in crop-yield, insurance, and adaptation studies makes cumulative degree-day exposure a standard empirical convention rather than a study-specific index.

Dairy research uses the same distinction with THI. N. R. St-Pierre et al. (2003) separately calculate heat-stress duration and `THI_Load`, the integral of THI above a threshold. J. M. Mbuthia et al. (2022) call the analogous hourly accumulation "cooling THI-days" and show that it adds information beyond conventional THI models. Our `THI69_Days` measure records duration: every hour at or above THI 69 contributes 1/24 of a stress day, irrespective of severity. `THI69_Load` is the corresponding intensity-duration measure and is mathematically closer to harmful or extreme degree days. Section 3.3 defines both constructions, and

Section 6.1 compares their estimates. The preferred results agree in sign across the two measures, showing robustness to emphasizing exposure duration or cumulative heat load; their coefficients are not directly comparable because their units and transformations differ.

2.3 Damage estimates

At the national scale, heat stress imposes well-documented costs on U.S. dairy production. N. R. St-Pierre et al. (2003) combine animal-response functions with long weather histories and estimate average annual dairy losses near $897 million without heat abatement. N. Key et al. (2014) use operation-level survey data merged with finely scaled climate records and find that heat stress reduced the average farm's 2010 production value by roughly $39,000, with losses concentrated in southern states. G. Mauger et al. (2014) project present-day national yield losses of 1.9 percent relative to a no-stress baseline, rising to 6.3 percent by the end of the century under continued warming. Regional production-frontier studies show that cooling systems can offset part of these losses, though the response varies across climate zones and production systems (D. Mukherjee et al., 2012; R. A. Lopez et al., 2022). At the farm level, J. Hutchins et al. (2025) find losses concentrated among small operations facing the greatest barriers to adaptation investment.

Heat stress also degrades milk composition, which matters independently for revenue under component-based pricing. O. Ravagnolo et al. (2000) show that fat and protein yields begin to decline once THI exceeds 72, at rates of approximately 0.012 and 0.009 kg per THI unit, respectively. W. Vroege et al. (2023) study both quantity and quality jointly in a European farm panel and find that composition losses amplify the revenue consequences of volume reduction. J. Hutchins et al. (2025) similarly report that quality effects on farm income exceed the volume effect alone. Most directly relevant to our setting, J. Prieto et al. (2026) link test-day records for roughly 6.5 million U.S. cows to high-resolution gridded THI data and show that fat and protein content begin declining at lower THI thresholds than yield, extending losses well beyond summer. When priced through Federal Milk Marketing Order component rates, total revenue damage is approximately twice the yield-only estimate, around $1.65 billion per year, with no detectable adaptation across farm size, region, or time period within their sample. A prior work at our geographic unit is X. Fan and J. Ma (2024), who use state-year panel data of weather and milk productivity in the US. Using spatial-error

and spatial-Durbin models, they find an inverted-U relationship between heat stress and milk production per cow. Their analysis highlights that standard two-way fixed-effects residuals carry significant spatial autocorrelation, which can bias estimates and inflate standard errors. Our approach addresses nonlinearity and confounding through a flexible polynomial basis expansion within the ADML debiasing framework rather than through an assumption of explicit spatial error structure.

Cold stress has received less attention in economic studies, but cold effects appear in large observational panels. M. Gisbert-Queral et al. (2021) find that U.S. state-level yields fall by approximately 6 percent under extreme cold, a magnitude rivaling heat losses at the national scale. E. Choi et al. (2024) corroborate this finding across multiple thermal indices and show that the yield decline at the cold tail of the THI distribution is gradual and continuous rather than threshold-shaped. While controlled physiological experiments and seasonal studies indicate that cold exposure elevates milk fat concentration (e.g., J. M. L. Heck et al., 2009; I. J. Salfer et al., 2019), large-scale observational panel studies have focused almost exclusively on heat stress and milk volume (e.g., N. R. St-Pierre et al., 2003; N. Key et al., 2014; X. Fan and J. Ma, 2024; J. Prieto et al., 2026), leaving observational cold-stress impacts on milk composition unexamined in the econometric literature. Together, these findings motivate our two-sided exposure design.

3 Data and exposure construction

3.1 Outcomes

Monthly milk production per cow and milk fat content come from USDA NASS QuickStats (USDA National Agricultural Statistics Service (NASS)). The balanced production panel covers 21 major dairy states from 2003 through 2025 and represents roughly 90 percent of U.S. milk output. The quantity outcome is log monthly pounds of milk per cow (`lmpc`). The composition outcome is milk fat content (`fc`) in percentage points.

Analyzing both margins is essential for understanding the full economic and biological consequences of climate stress on the dairy sector. Milk volume per cow captures physical output and animal energetic efficiency, directly determining aggregate raw milk supply and baseline farm cash receipts. Milk fat content, by contrast, dictates industrial manufacturing

yields for high-value dairy products (such as butter, cheese, and milk powder) and represents the primary revenue determinant under Federal Milk Marketing Order (FMMO) component pricing formulas. While volume and composition are co-primary outcomes in our empirical estimation, our aggregate monetary valuation and long-run adaptation analysis focus on `lmpc` because federal component price benchmarks and uniform valuation schedules are not available across the full historical panel.

USDA records omit milk per cow for March through June 2013. We impute those observations with the corresponding state-calendar-month average from 2012 and 2014 to preserve the longitudinal continuity required for first differences and distributed lags across the full 21-state panel. Fat content has 48 initial missing observations spanning January 2003 through December 2004 for Colorado (24 months), and January 2004 through December 2004 for Kansas (12 months) and Oregon (12 months). We leave those values missing and exclude the affected first differences rather than impute composition. We impute the brief 4-month gap in milk per cow because it was an isolated administrative suspension in an otherwise unbroken 23-year panel, where omitting observations would rupture first differences and distributed lags across all 21 states; in contrast, the missing fat content observations are initial prefix blocks at the start of the sample prior to NASS component survey expansion, allowing clean sample truncation without interior longitudinal rupture or artificial composition imputation error.

3.2 Hourly weather and THI

ERA5-Land supplies hourly 2-meter air temperature, dew point, precipitation, 10-meter wind components, and surface solar radiation. We convert temperature and dew point from Kelvin to degrees Celsius and calculate relative humidity RH using the standard Magnus approximation. We then apply the Yeck THI formula used in the project weather-construction code:

$$THI = 1.8T + 32 - (0.55 - 0.0055RH)(1.8T - 26),$$

where T is 2-meter air temperature in degrees Celsius and RH is relative humidity in percent.

The weather data construction proceeds in three sequential stages: (1) hourly THI index calculation at the grid-cell level, (2) temporal aggregation into 24-hour-equivalent exposure days within each county-month, and (3) spatial aggregation from counties to states using time-varying milk-cow inventory weights. Monthly precipitation is the sum of hourly millimeters; wind speed and solar radiation are monthly means.

3.3 Climate stress measurements

Climate-economics research shows why nonlinear weather responses and spatial aggregation matter. Degree-day and bin designs recover damage at temperature tails that mean-temperature regressions can obscure (W. Schlenker and M. J. Roberts, 2009; S. Hsiang, 2016). Aggregation can also attenuate extremes when weather is averaged over land areas that do not contain dairy cattle or milk production. Crop studies address this aggregation bias by weighting weather exposure toward cropland or planted area (W. Schlenker and M. J. Roberts, 2009; O. Deschênes and M. Greenstone, 2007; M. Burke and K. Emerick, 2016); we adapt a similar approach for dairy by weighting toward milk-cow inventory.

The main model uses duration-day measures. For a county-month with H hourly observations, we define

$$\texttt{THI39_Days} = \frac{1}{24}\sum_{h=1}^{H}\mathbf{1}(THI_h < 39), \qquad \texttt{THI69_Days} = \frac{1}{24}\sum_{h=1}^{H}\mathbf{1}(THI_h \geq 69).$$

Thus, `THI39_Days` records 24-hour-equivalent exposure below THI 39 and `THI69_Days` records equivalent exposure at or above THI 69, where each stress day represents 24 non-consecutive stressful hours within the county-month without assigning extra weight to more severe hours. THI from 39 through 69 is the omitted thermoneutral reference. The thresholds follow the lower-tail U.S. evidence and the lower heat threshold observed for modern high-producing cows (M. Gisbert-Queral et al., 2021; E. Choi et al., 2024; T. M. McWhorter et al., 2023).

To account for thermal intensity beyond binary thresholds, the empirical model incorporates higher-order polynomial expansions (quadratic and cubic terms in `THI69_Days` and `THI39_Days`) and interactions with precipitation. As a direct intensity-weighted alternative,

we also construct intensity-duration load measures that accumulate each hourly deviation beyond the same thresholds:

$$\texttt{THI39_Load} = \sum_{h=1}^{H} \max(0, 39 - THI_h), \qquad \texttt{THI69_Load} = \sum_{h=1}^{H} \max(0, THI_h - 69).$$

An hour farther beyond the threshold therefore contributes proportionally more than a mildly stressful hour, measuring cumulative thermal load in THI-hours per county-month. Because the raw loads are highly skewed and can equal zero, estimation uses shifted logs:

$$\texttt{ln_THI39_Load} = \ln(0.001 + \texttt{THI39_Load}), \qquad \texttt{ln_THI69_Load} = \ln(0.001 + \texttt{THI69_Load}).$$

The 0.001 shift retains zero-load months; the logarithm compresses the upper tail and gives the estimated marginal effects an approximate elasticity interpretation away from zero. The cleaned data retain `CLI39`, `HLI69`, `LCLI39`, and `LHLI69` as exact legacy aliases for these four measures.

This two-bin design is a deliberate power-versus-resolution choice. A seven-bin heat specification leaves the most extreme cells sparse and produces imprecise ADML estimates in a 21-state panel. We retain broad tails and recover within-tail nonlinearity through cubic terms and THI-by-precipitation interactions. Section 6.1 reports the duration-day and intensity-duration estimates together.

3.4 Dairy-cow-weighted spatial aggregation

We construct exposure at the county level and aggregate it with time-varying milk-cow inventory weights. This avoids treating sparsely populated counties as equally relevant to dairy production. The resulting state-month treatment therefore measures weather where cows are located, not the average condition over a state's land area. This geographic alignment is especially important in large states with concentrated dairy regions.

We first aggregate ERA5-Land grid cells to counties. County exposures are then averaged within each state-month using county milk-cow inventories from the 2007, 2012, 2017, and

2022 agricultural censuses. Each sample year uses its nearest designated census anchor: 2007 for 2003-2009, 2012 for 2010-2014, 2017 for 2015-2019, and 2022 for 2020-2025. Counties without supported cow inventory receive zero weight. The measure therefore follows the within-state location of dairy herds.

3.5 Descriptive statistics

3.5.1 Summary statistics

Table 1 reports descriptive statistics for the main variables used in the state-month analysis dataset before first-differencing and lag construction. The main ADML productivity model uses a smaller effective estimation sample after requiring lagged weather histories and first differences. Fat content has 48 fewer observations because Colorado, Kansas, and Oregon have missing fat-content values at the beginning of the sample.

Table 1. Summary statistics for the main state-month modeling variables.

Variable	Unit	N	Mean	SD	Min	P25	Median	P75	Max
`lmpc`	log lb/cow	5,796	7.500	0.122	6.928	7.415	7.507	7.596	7.783
`mpc`	lb/cow	5,796	1,821	218	1,020	1,660	1,820	1,990	2,400
`fc`	percent	5,748	3.824	0.241	3.310	3.650	3.780	3.950	4.760
`THI69_Days`	days/month	5,796	6.223	8.473	0.000	0.000	1.345	10.613	31.000
`THI39_Days`	days/month	5,796	6.601	9.191	0.000	0.000	1.004	11.609	30.997
`precip_mm`	mm/month	5,796	71.682	49.265	0.023	32.832	65.613	100.883	373.098

Note: `lmpc` is log milk production per cow; `mpc` is monthly milk production per cow in pounds; `fc` is milk fat content in percent; `THI69_Days` is days with THI ≥ 69; `THI39_Days` is days with THI < 39; `precip_mm` is monthly precipitation in millimeters. Sample covers 21 major dairy states from 2003 through 2025.

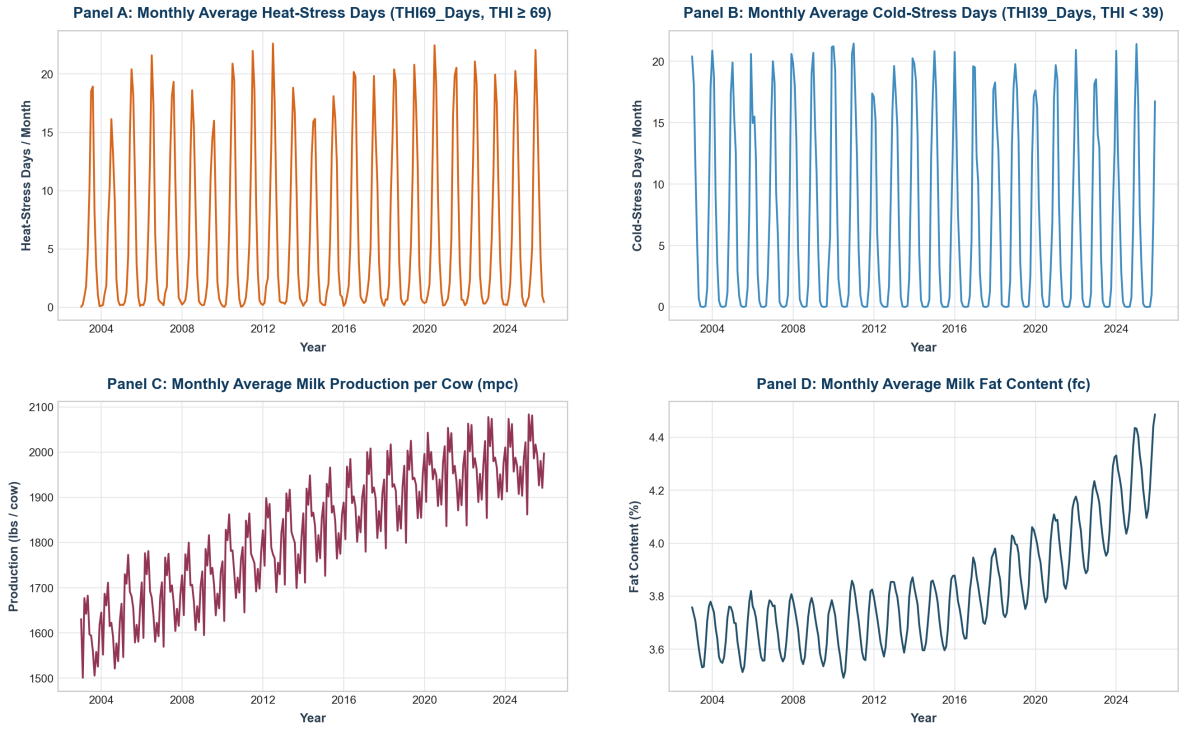
3.5.2 Temporal and spatial patterns

Figure 1 illustrates the aggregate monthly dynamics of thermal stress and dairy productivity across the 21 study states from 2003 through 2025. Panels A and B display the pronounced seasonal cycles of heat stress (`THI69_Days`) and cold stress (`THI39_Days`), respectively. Heat exposure peaks sharply in July and August, while cold exposure concentrates heavily in December through February.

Panels C and D track the primary dairy productivity outcomes. Monthly milk production per cow (`mpc`, Panel C) shows a combination of strong intra-annual seasonality (peaking during spring months and bottoming out during summer heat) and a continuous long-term

upward trend resulting from genetic selection, nutritional optimization, and technological adoption. Milk fat content (`fc`, Panel D) demonstrates pronounced seasonal oscillations (dropping during summer heat stress and concentrating during cold weather) along with a structural increase in average butterfat levels in later years reflecting component-pricing incentives.

Figure 1. Monthly temporal variations in weather stress exposure and dairy productivity (2003-2025).



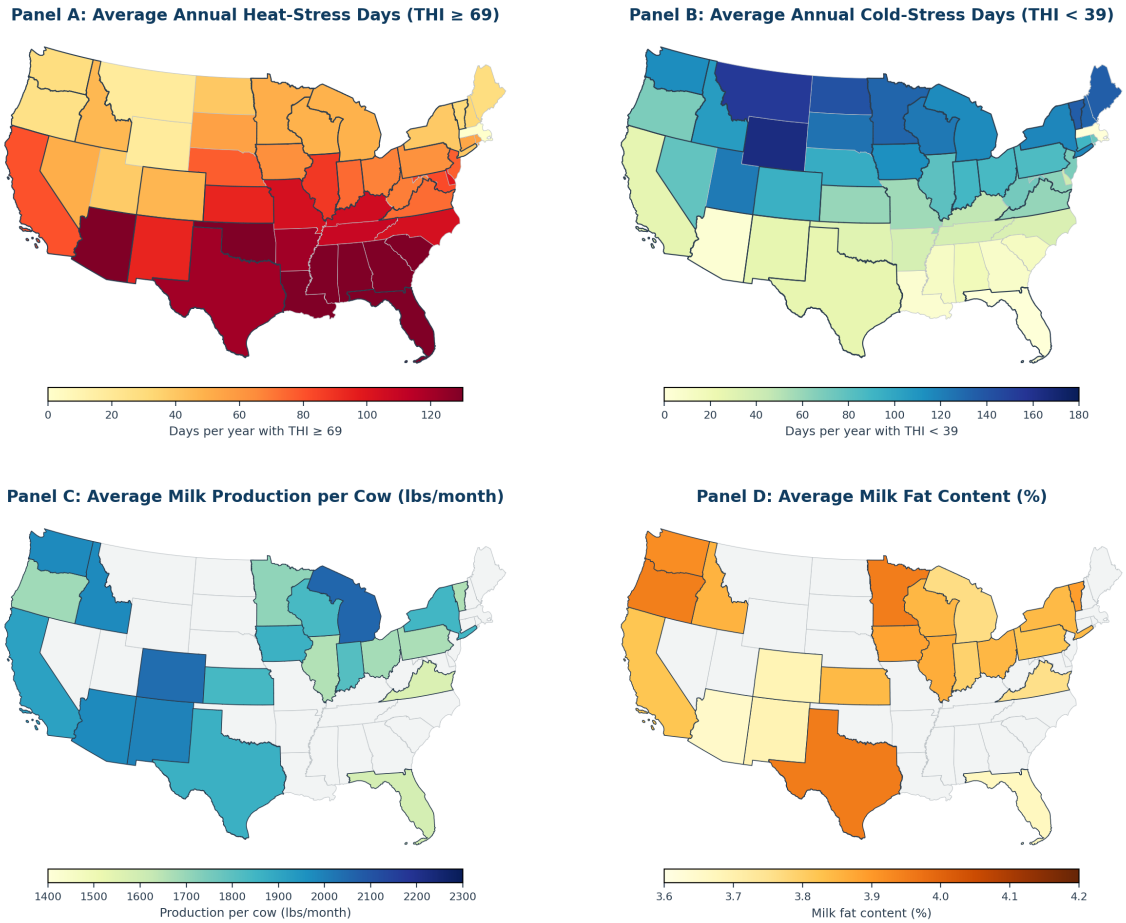
Note: Panel A plots the monthly average heat-stress days ($THI69_Days$, THI $\geq$ 69); Panel B plots the monthly average cold-stress days ($THI39_Days$, THI $<$ 39); Panel C plots monthly average milk production per cow (mpc in pounds); Panel D plots monthly average milk fat content (fc in percent). All series are monthly averages across the 21 primary study states over the 2003-2025 sample period.

To complement these aggregate temporal patterns, Figure 2 maps the spatial distribution of long-run weather exposure and productivity across the United States. Panel A displays average annual heat-stress days, highlighting that heat exposure is concentrated in Southern and Southwestern states (e.g., Texas, Florida, Arizona), where dairy herds experience over 100 heat-stress days per year. Conversely, Panel B shows that cold stress is heavily concentrated in Northern and Midwestern states (e.g., Wisconsin, Minnesota, Idaho).

Panels C and D map the spatial patterns of dairy productivity across the 21 major dairy-

producing study states. As shown in Panel C, average per-cow milk yields are highest in the West and Southwest (e.g., Colorado, California, Idaho), where large-scale, highly managed dry-lot and free-stall operations predominate. Panel D demonstrates that milk fat content is highest in Northern states (e.g., Oregon, Minnesota, Vermont), where cooler climates and higher concentrations of high-butterfat breeds (such as Jerseys) support higher milk solids.

Figure 2. Spatial distribution of average annual weather stress and dairy productivity (2003-2025).



Note: Panel A maps average annual heat-stress days (THI $\geq$ 69); Panel B maps average annual cold-stress days (THI $<$ 39); Panel C maps long-run average monthly milk production per cow (lbs/month); Panel D maps long-run average milk fat content (percent). The 21 major dairy-producing study states are highlighted with dark outlines. Non-study states in Panels C and D are shown in light grey.

4 Econometric framework

4.1 Motivation and why ADML

Estimating the productivity and composition impacts of climate stress in observational panels presents two central econometric challenges: nonlinear response surfaces with high-dimensional confounding, and regularized estimation bias.

The dominant empirical approach in climate economics is linear two-way fixed-effects (TWFE) regression. When the true biological dose-response exhibits sharp thresholds, polynomial curvature, and moisture compounding effects, standard linear TWFE estimates are biased for the average partial derivative (W. Schlenker and M. J. Roberts, 2009; S. Klosin and M. Vilgalys, 2025). Weather shocks provide credible exogenous variation because short-run weather is not chosen by producers, but forcing a linear functional form onto a nonlinear biological response assigns curvature and interaction effects to the error term, generating systematic misspecification bias.

Nonparametric machine learning can approximate complex nonlinear response surfaces flexibly, but regularization introduces a second problem: shrinkage penalties bias estimated derivative functionals toward zero. Automatic Debiased Machine Learning (ADML) resolves both issues simultaneously. It pairs a flexible ML-estimated regression function with an orthogonal Riesz-representer correction that makes the target causal parameter first-order insensitive to errors in nuisance function estimation (V. Chernozhukov et al., 2022; S. Klosin and M. Vilgalys, 2025). For continuous climate treatments, the Riesz representer supplies correction weights directly, avoiding the numerical instability associated with inverting small estimated treatment densities in classical propensity-score debiasing.

Because theoretical debiasing guarantees hold asymptotically, we calibrate semi-synthetic Monte Carlo simulations directly to our 21-state empirical weather panel in Section 5 to evaluate finite-sample point accuracy and interval coverage before interpreting the empirical estimates.

4.2 Panel model, identification, and estimand

Let y_{it} denote the outcome in state i and month t. For the quantity margin, y_{it} is log monthly milk production per cow (`lmpc`); for the composition margin, it is milk fat content (`fc`). The semiparametric panel model is

$$y_{it} = a_i + \gamma_t(D_{it}, X_{it}) + \varepsilon_{it}, \tag{1}$$

where a_i is an unobserved state fixed effect, D_{it} is the continuous thermal treatment vector of interest, X_{it} is a vector of time-varying covariates, and ε_{it} satisfies strict exogeneity conditional on the full history of treatments and covariates (S. Klosin and M. Vilgalys, 2025):

$$\mathbb{E}[\varepsilon_{it} \mid D_{i1}, \dots, D_{iT}, X_{i1}, \dots, X_{iT}, a_i] = 0.$$

Because short-run weather is determined by meteorological processes rather than producer choice, strict exogeneity is more credible than in management or policy evaluations. We also assume a common underlying response function $\gamma_t(\cdot)$ across dairy cows conditional on state fixed effects and flexible covariates. This assumption enables pooling weather shocks across all 21 major dairy states to identify the semiparametric response surface, while state fixed effects a_i absorb persistent differences in herd genetics, baseline management, housing infrastructure, and regional endowments.

The treatment vector D_{it} captures monthly thermal exposure across current and lagged months. The covariate vector X_{it} includes monthly precipitation, temporal fixed effects, and within-lag interactions between precipitation and thermal stress bins to account for humidity and moisture compounding.

The parameter of interest is the lag-ℓ Average Partial Derivative (APD),

$$\tau_\ell = \mathbb{E}\left[\frac{\partial \gamma_t(D_{it}, X_{it})}{\partial D_{i,t-\ell}}\right], \tag{2}$$

which measures the expected change in the selected outcome associated with an additional day of thermal exposure at lag ℓ. The headline parameter is the cumulative distributed-lag effect over a five-month response window ($\ell \in \{0, 1, 2, 3, 4\}$):

$$\delta_{\text{cum}} = \sum_{\ell=0}^{4} \tau_\ell.$$

For `lmpc`, δ_{cum} represents a log-point change, approximately equal to the percentage change in monthly milk production per cow per additional stress day. For `fc`, it represents

a percentage-point change in milk fat concentration.

Importantly, the APD parameter identifies the expected marginal effect of continuous, multi-day exposure across the observed support. We do not estimate an Average Treatment Effect on the Treated (ATT), because observational weather exposure is continuous, distributed over time, and experienced across all states rather than functioning as a discrete binary intervention with an untreated control group.

4.3 Specification and implementation

Following S. Klosin and M. Vilgalys (2025), we eliminate additive state fixed effects a_i by first-differencing:

$$\Delta y_{it} = y_{it} - y_{i,t-1} = \Delta\gamma(D_{it}, X_{it}) + \Delta\varepsilon_{it}, \tag{3}$$

where

$$\Delta\gamma(D_{it}, X_{it}) = \gamma_t(D_{it}, X_{it}) - \gamma_{t-1}(D_{i,t-1}, X_{i,t-1}). \tag{4}$$

First-differencing removes static unit fixed effects exactly without requiring sparsity assumptions on unit intercepts.

A crucial methodological requirement is that the distributed-lag structure must be specified in levels prior to differencing. While first-differencing removes time-invariant unit confounders, it does not substitute for substantive distributed-lag modeling. Biological recovery, delayed lactation curve depression, and metabolic exhaustion operate across multiple months. Preserving multi-month distributed lags in levels before differencing ensures that dynamic carry-over effects are fully captured.

4.3.1 Climate exposure data and basis expansion

The primary thermal exposure measures are coarse duration-day bins: `THI39_Days` (days with THI $<$ 39) and `THI69_Days` (days with THI $\geq$ 69), with the thermoneutral interval THI $\in [39, 69)$ serving as the omitted baseline. Let $B_{i,t-\ell} = (\texttt{THI39_Days}_{i,t-\ell}, \texttt{THI69_Days}_{i,t-\ell})'$ denote the vector of coarse thermal exposures at lag $\ell \in \{0, 1, 2, 3, 4\}$, and let $P_{i,t-\ell}$ denote monthly precipitation at the same lag.

The flexible basis expansion is constructed in levels as:

$$X_{it}^{\mathrm{SR}} = \Big\{ \{B_{i,t-\ell}, B_{i,t-\ell}^2, B_{i,t-\ell}^3\}_{\ell=0}^4,\ \{P_{i,t-\ell}, P_{i,t-\ell}^2, P_{i,t-\ell}^3\}_{\ell=0}^4,\ \{B_{i,t-\ell} \otimes P_{i,t-\ell}\}_{\ell=0}^4 \Big\}, \quad (5)$$

where $\otimes$ denotes within-lag interactions between thermal-bin exposures and monthly precipitation.

This flexible polynomial and interaction specification serves as the preferred empirical model. Coarse THI bins preserve statistical power across the 21-state panel, while quadratic and cubic terms alongside THI-by-precipitation interactions capture nonlinear marginal responses and moisture compounding effects, avoiding the severe attenuation bias of linear models documented in our Monte Carlo simulations.

After constructing basis terms in levels, all features are first-differenced. Because the level specification incorporates four distributed lags $(t, t-1, t-2, t-3, t-4)$, first-differencing $(\Delta X_{i,t-4} = X_{i,t-4} - X_{i,t-5})$ requires historical weather data dating back five months $(t-5)$. This timing requirement defines the effective estimation sample and reinforces the need to keep complete state time series intact within cross-fitting folds.

4.3.2 Additional meteorological controls

The preferred main specification focuses on THI, precipitation, their polynomial expansions, and their within-lag interactions. In robustness checks, we also incorporate wind speed and solar radiation from ERA5-Land. Monthly wind speed is calculated from 10-meter neutral horizontal wind components ($\sqrt{u_{10}^2 + v_{10}^2}$ in m/s), and solar radiation is the monthly mean surface downward solar radiation flux in $\mathrm{W/m^2}$.

These controls capture distinct biophysical heat-exchange channels: wind enhances convective heat dissipation from the cow's skin and coat, alleviating thermal stress, whereas solar radiation imposes a direct radiative heat load on dairy cattle in pasture and open-lot housing systems.

4.3.3 Time fixed effects and unobserved national shocks

The baseline time fixed-effect specification includes separate year fixed effects and calendar month-of-year fixed effects. This additive specification absorbs long-run macroeconomic trends, technological progress, and nationwide seasonal cycles in dairy yield and composition,

while preserving identifying variation from differential weather shocks across states within a given year-month.

As a robustness check, we also estimate models with full month-by-year interaction fixed effects. This specification absorbs all common national shocks at the calendar-month level, so that identification relies entirely on idiosyncratic cross-state weather deviations within each month-year cell. Consistency of estimates across additive and interacted time fixed effects confirms that our results are not driven by residual seasonality or omitted national macroeconomic shocks.

4.3.4 Cross-fitting and state-clustered inference

Cross-fitting is central to ADML estimation to prevent overfitting and sample-reuse bias. Following S. Klosin and M. Vilgalys (2025), cross-fitting folds are blocked at the state level: all observations from a given state remain within the same fold so that within-unit temporal dependence does not leak across training and evaluation splits.

Analytical standard errors and Webb six-point wild-cluster-bootstrap intervals are clustered by state (A. C. Cameron et al., 2008; M. D. Webb, 2023). Because Monte Carlo simulations reveal undercoverage under 21 clusters in finite samples, confidence intervals are interpreted as indicative bounds of uncertainty rather than exact finite-sample probability statements.

4.3.5 Estimator suite and benchmark comparisons

To evaluate econometric robustness and illustrate the role of flexible nuisance modeling and debiasing, we compare five distinct estimators across the same state-month panel:

1. **Linear OLS**: The standard linear two-way fixed-effects specification widely used in agricultural climate literature, estimating constant marginal effects across the five-month lag window.
2. **Full OLS**: OLS estimation using the full polynomial basis expansion and precipitation interactions in Equation (5).
3. **Plug-in Machine Learning (LASSO and Neural Network)**: Flexible nonparametric learners (cross-validated LASSO and multi-layer perceptron neural networks) that estimate the regression function $\gamma_t(\cdot)$ but evaluate derivative functionals directly

without Riesz representer debiasing.

4. **Automatic Debiased Machine Learning (ADML-LASSO and ADML-NNET)**: The preferred estimators, which combine regularized ML nuisance function estimation with Riesz representer orthogonal score debiasing to eliminate regularization bias.

4.4 Year-specific heat effects and temporal heterogeneity

The pooled empirical estimate measures the average short-run marginal damage of an additional `THI69_Days` day over the entire 2003-2025 period. However, this full-sample average may mask adaptation or technological changes over time. If dairy producers have adopted cooling infrastructure (e.g., cross-ventilation, evaporative misters, shade structures) or optimized ration management, the marginal damage per heat-stress day should attenuate in later years.

A standard fixed-effects regression analogue interacts annual heat exposure with calendar-year indicators:

$$y_{it} = a_i + \lambda_t + W_{it}'\theta + \sum_{\ell=0}^{4} \sum_{y} \beta_{y,\ell}^{H} H_{i,t-\ell} \mathbf{1}\{\text{year}(t) = y\} + u_{it}, \tag{6}$$

where H_{it} is `THI69_Days`, a_i is a state fixed effect, λ_t represents time controls, W_{it} contains baseline controls and flexible basis terms, and $\beta_{y,\ell}^{H}$ is the lag-ℓ heat effect in year y. The year-specific cumulative heat effect is:

$$\beta_{y,\text{cum}}^{H} = \sum_{\ell=0}^{4} \beta_{y,\ell}^{H}.$$

In the ADML framework, temporal heterogeneity is evaluated using observation-level cumulative orthogonal scores from the preferred NNET ADML specification. For each state-month observation, define the cumulative `THI69_Days` score:

$$\hat{\psi}_{it,\text{cum}}^{H} = \sum_{\ell=0}^{4} \hat{\psi}_{\ell,it}^{H}, \tag{7}$$

where H denotes `THI69_Days`. The year-specific heat effect $\hat{\tau}_y^H$ is the calendar-year average of these cumulative scores:

$$\hat{\tau}_y^H = \frac{1}{N_y} \sum_{(i,t):\text{year}(t)=y} \hat{\psi}_{it,\text{cum}}^H. \tag{8}$$

Econometrically, $\hat{\tau}_y^H$ estimates the average marginal effect of heat stress for observations in calendar year y, summed over lags 0 through 4. The year-specific decomposition is analogous to dynamic or event-study summaries of treatment effects, but the estimand remains an average partial derivative for a continuous weather treatment rather than a binary treatment effect (B. Callaway et al., 2024; S. Klosin and M. Vilgalys, 2025). The resulting plot should therefore be read as a descriptive decomposition of the ADML heat effect over time, not as proof by itself that a specific adaptation mechanism changed in a specific year.

5 Estimator validation: semi-synthetic Monte Carlo simulation

5.1 Simulation design

The estimator comparison is supported by a semi-synthetic Monte Carlo simulation. The purpose of the simulation is not to reproduce the full empirical data-generating process, which is unknown, but to evaluate whether the estimators recover known average partial derivatives in a panel with the same state-month weather design, lag structure, fixed-effect structure, and sample size as the empirical application.

The simulation keeps the observed dairy weather panel fixed and replaces the outcome with a generated log milk-production outcome. The baseline component preserves state-level mean differences, year effects, month-of-year seasonality, and serially correlated state-level shocks calibrated to the residual scale of the observed outcome. The structural weather component is known by construction, so the true current, lagged, and cumulative average partial derivatives for `THI39_Days` and `THI69_Days` are known in each replication.

The main simulation uses a smooth nonlinear DGP with nonlinear strength 3.0, representing a moderate degree of curvature calibrated so that true cumulative marginal damages fall within the empirical range of observed heat effects. This DGP combines threshold-like heat and cold responses, saturating nonlinear terms, seasonal interactions, and precipitation interactions. It is deliberately not exactly spanned by the polynomial basis expansion, so the comparison does not mechanically favor Full OLS or Lasso.

The Monte Carlo comparison includes six estimator arms:

- Linear OLS, using linear THI-bin, precipitation, and time-control terms.
- Full OLS, using the full polynomial and interaction basis expansion.
- LASSO no Riesz, using cross-fitted plug-in derivatives without Riesz correction.
- LASSO ADML, using regular LASSO plus the Riesz correction.
- NNET no Riesz, using cross-fitted neural-network derivatives.
- NNET ADML, using the neural-network learner plus the Riesz correction.

The plug-in arms isolate the role of the Riesz correction for each machine learner. The estimator ranking is therefore interpreted conditionally on the DGP being studied. NNET ADML is designated as preferred because it has the lowest RMSE under the main smooth nonlinear dairy design, not because ADML is assumed to dominate simpler estimators in every data-generating process.

Each replication reports bias, root mean squared error, empirical standard deviation, mean analytic standard error, and confidence-interval coverage for each lag-specific and cumulative estimand. In addition to analytic state-clustered intervals, the simulation reports score-based wild cluster bootstrap intervals using state-level Webb six-point weights. The bootstrap follows the cluster bootstrap logic of A. C. Cameron et al. (2008) and the M. D. Webb (2023) small-cluster weight distribution. It is applied directly to the observation-level estimator scores, so it recalibrates inference without refitting the first-stage or Riesz nuisance functions in each bootstrap draw.

The main simulation summary focuses on the cumulative `THI69_Days` estimand because this is the primary heat-stress parameter. The same diagnostics are retained for `THI39_Days` and for individual lags to check whether any estimator performs well only after lag aggregation.

5.2 Simulation results

We evaluate estimator performance across 1,000 replications of the main smooth nonlinear DGP, where the known true cumulative `THI69_Days` average partial derivative is $\tau_0 = -0.003189$. Table 2 reports the mean estimate, bias, empirical standard deviation, and root mean squared error (RMSE) for each estimator, and Figure 3 visualizes the corresponding error distributions. In short, NNET ADML achieves the lowest overall RMSE

by balancing the bias reduction of orthogonal Riesz debiasing with the variance-reducing properties of neural-network regularization.

Both the bias and variance comparisons illustrate the distinct trade-offs across estimators under a nonlinear response surface. First, regarding bias, the orthogonal Riesz correction substantially reduces first-stage estimation error for the neural network learner: NNET ADML reduces the plug-in bias of NNET no Riesz by over 31 percent, from 0.000813 to 0.000557. In contrast, LASSO without and with the Riesz correction exhibits substantial shrinkage bias (-0.00106 and -0.00155, respectively), as L_1 penalization on a high-dimensional polynomial basis struggles to span the smooth nonlinear response surface. Linear OLS displays low bias in this specific smooth calibration (-0.000277) but lacks the flexibility to adapt to more complex interaction surfaces, while Full OLS attempts to capture curvature through raw polynomial terms but suffers from severe overparameterization.

Second, we evaluate total estimation error via RMSE, which incorporates both bias and variance. NNET ADML achieves the lowest RMSE (0.000634) among all six estimators evaluated. Due to neural-network regularization and cross-fitting, NNET ADML maintains a low standard deviation (0.000302)—less than half that of Linear OLS (0.000629) and nearly an order of magnitude lower than Full OLS (0.00228). As a consequence, NNET ADML achieves an RMSE roughly half that of LASSO no Riesz (0.00113), less than half that of LASSO ADML (0.00160), and less than one-third that of Full OLS (0.00238).

Figure 3 visualizes these performance differences through the simulated error distributions (defined as the estimated minus the true cumulative derivative). The error distribution for NNET ADML is tightly centered around zero with minimal dispersion. In contrast, Full OLS exhibits a wide, heavy-tailed distribution reflecting the variance penalties of multicollinear basis expansion in a 21-cluster panel, while the LASSO estimators display pronounced leftward shifts driven by regularization shrinkage.

These simulation results demonstrate the practical value of panel ADML for applied agricultural climate economics. NNET ADML provides an effective empirical alternative that avoids both the functional-form rigidity of Linear OLS and the variance inflation of Full OLS, while successfully mitigating first-stage plug-in bias. In the empirical analyses that follow, we designate NNET ADML as our preferred estimator for point estimation and

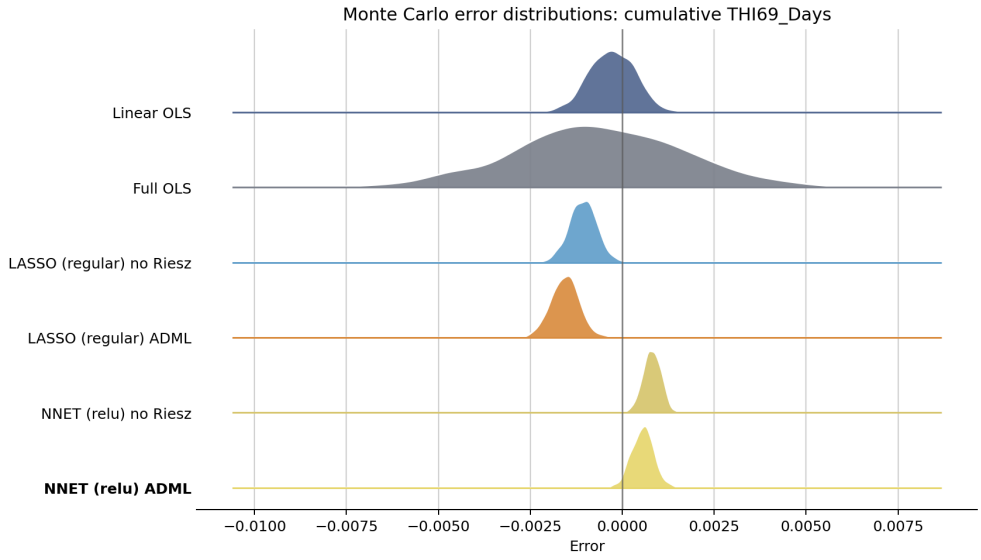
examine the stability of response magnitudes across the full set of comparison estimators. In the valuation, the Webb endpoints define sensitivity bands rather than exact 95 percent confidence statements.

Table 2. Monte Carlo performance for the cumulative `THI69_Days` average partial derivative under the main smooth nonlinear DGP.

Estimator	Mean estimate	Bias	Std. dev.	RMSE
Linear OLS	-0.00347	-0.000277	0.000629	0.000687
Full OLS	-0.00387	-0.000681	0.00228	0.00238
LASSO no Riesz	-0.00425	-0.00106	0.000375	0.00113
LASSO ADML	-0.00474	-0.00155	0.000386	0.00160
NNET no Riesz	-0.00238	0.000813	0.000250	0.000851
NNET ADML	-0.00263	0.000557	0.000302	0.000634

Note: The table reports bias, standard deviation, and root mean squared error (RMSE) for the cumulative `THI69_Days` average partial derivative based on 1,000 replications. The true cumulative target is -0.00319. `THI69_Days` represents Heat-stress exposure days (Temperature Humidity Index THI $\geq$ 69).

Figure 3. Monte Carlo error distributions for the cumulative `THI69_Days` estimand.



Note: Errors are estimated effects minus the known true average partial derivative.

6 Results

The empirical comparison holds a common treatment and control design fixed across six estimators. The `lmpc` analysis uses 5,691 first-differenced state-month observations from 21 states; the `fc` analysis uses 5,643 observations after excluding 48 initially missing fat-content

records. The headline estimand is the cumulative marginal effect of `THI69_Days` over the current month and four monthly lags. Cold exposure is measured by `THI39_Days`, and the THI 39-69 range is the omitted thermoneutral category.

At each lag, the basis expansion includes the heat- and cold-stress measures, precipitation, linear through cubic terms in each weather variable, and within-lag THI-by-precipitation interactions. The baseline absorbs year and month-of-year fixed effects. NNET ADML is the preferred estimator, combining state-blocked cross-fitting with the Riesz correction; Linear OLS, Full OLS, LASSO without and with the Riesz correction, and the NNET plug-in estimator provide transparent functional-form and debiasing comparisons.

The robustness exercises preserve this core design while replacing degree-day exposure with shifted-log heat- and cold-load indexes, adding wind speed and solar radiation, or using full month-year fixed effects.

6.1 Quantity margin: milk production per cow

Because the outcome is log milk production per cow, the preferred NNET ADML heat estimate implies that one additional `THI69_Days` day is associated with an approximately 0.319 percent decline in monthly milk production per cow over the five-month response window. The Riesz correction changes the neural-network plug-in estimate only modestly in the observed data, from -0.002783 to -0.003189. All six estimators imply a negative cumulative heat effect, while the magnitude remains sensitive to functional form. Cold-stress magnitudes are more estimator-sensitive and are treated as supporting evidence. The larger NNET ADML standard error is substantively important. For cumulative `THI69_Days`, the analytic standard error increases from 0.000226 for NNET no Riesz to 0.000773 for NNET ADML. This should not be interpreted mechanically as worse performance: the plug-in interval ignores first-stage estimation error, whereas the orthogonal score incorporates the Riesz-weighted outcome residual. Consistent with that distinction, the main Monte Carlo shows severe undercoverage for NNET no Riesz and substantially improved, though still imperfect, coverage for NNET ADML.

Table 3. Empirical estimates of cumulative cold- and heat-stress effects on log milk production per cow (`lmpc`) using duration-day and shifted-log intensity-duration measures.

Estimator	Cold effect (SE)	Heat effect (SE)
Panel A. Duration days		
Linear OLS	-0.00431 (0.00147)	-0.00301 (0.00117)
Full OLS	-0.00619 (0.00216)	-0.00951 (0.00150)
LASSO no Riesz	-0.00136 (0.0000880)	-0.00312 (0.000294)
LASSO ADML	-0.00136 (0.000272)	-0.00355 (0.000888)
NNET no Riesz	-0.000758 (0.0000600)	-0.00278 (0.000226)
NNET ADML	**-0.000782 (0.000293)**	**-0.00319 (0.000773)**
Panel B. Intensity load (log)		
Linear OLS	-0.00133 (0.00289)	0.00994 (0.00252)
Full OLS	-0.0138 (0.00211)	-0.0169 (0.00396)
LASSO no Riesz	-0.0103 (0.00115)	-0.0143 (0.00101)
LASSO ADML	-0.0136 (0.00180)	-0.0195 (0.00629)
NNET no Riesz	-0.00509 (0.000722)	-0.00748 (0.000802)
NNET ADML	**-0.00803 (0.00171)**	**-0.0138 (0.00564)**

Note: Reported effects are cumulative average partial derivatives over lags 0-4; analytic state-clustered standard errors are in parentheses. Both panels use 5,691 first-differenced observations from 21 states, year and month-of-year fixed effects, precipitation controls, five state-level cross-fitting folds, and the same learner settings. Panel A uses equivalent days below THI 39 and at or above THI 69. Panel B uses the shifted-log intensity-duration measures defined in Section 3.3. For preferred NNET ADML, the Panel B Webb 95 percent intervals are [-0.0115, -0.00477] for cold load and [-0.0240, -0.00383] for heat load.

Panel B of Table 3 asks whether the quantity result survives when stressful hours receive more weight as their THI moves farther beyond the threshold. The preferred NNET ADML estimate remains negative for both tails, and both Webb intervals exclude zero. For heat, five of the six estimators are negative; the exception is Linear OLS, reinforcing the importance of allowing a nonlinear weather response. Because the treatments are logs, the preferred coefficients are approximate elasticities with respect to `THI_Load`. A 1 percent increase in heat load is associated with an approximately 0.0138 percent reduction in milk production per cow over the five-month response window. This estimate is not numerically comparable to the effect per `THI69_Days` day, but its sign supports the main conclusion when exposure is weighted by intensity as well as duration. The ADML estimates also differ from their no-Riesz counterparts, indicating that the orthogonal corrections are active after the log transformation.

6.2 Composition margin: milk fat content

For fat content, the preferred cumulative `THI69_Days` estimate is -0.002456, with analytic standard error 0.000921 and Webb wild-cluster-bootstrap 95 percent interval [-0.004103, -0.000750]. This means that one additional heat-stress day is associated with a 0.00246 percentage-point decline in milk fat content over lags 0-4. Relative to the sample mean fat content of about 3.82 percent, the estimate is approximately a 0.064 percent decline in the fat content level. The direction is consistent with heat stress reducing both milk quantity and milk composition, but the fat-content effect is measured on a different scale and should not be interpreted as a percent change in output. The preferred `THI39_Days` estimate for fat content is positive, 0.004430 (SE 0.000287), suggesting a different cold-stress composition response than the negative cold-stress estimate for log production per cow.

Table 4. Empirical estimates of cumulative cold- and heat-stress effects on milk fat content (`fc`) using duration-day and shifted-log intensity-duration measures.

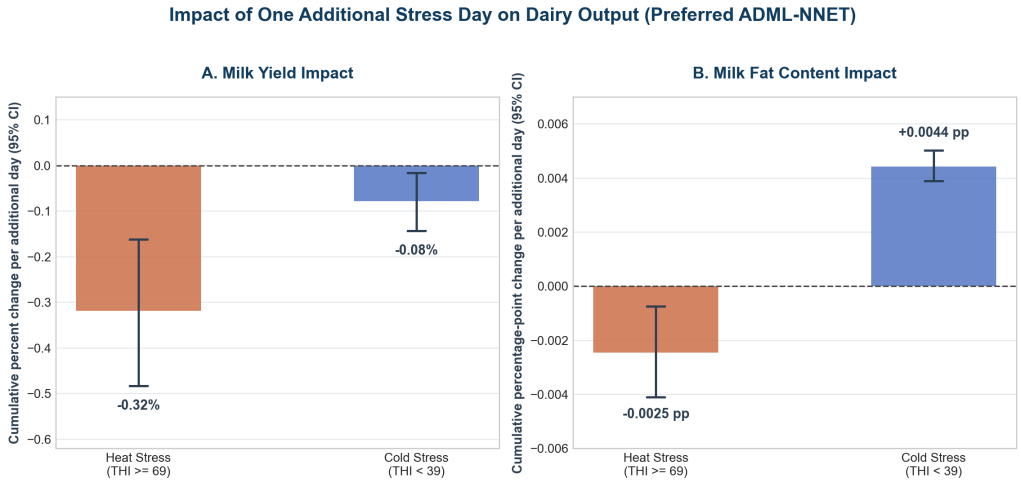
Estimator	Cold effect (SE)	Heat effect (SE)
Panel A. Duration days		
Linear OLS	0.00749 (0.00110)	-0.00579 (0.000829)
Full OLS	0.0160 (0.00293)	-0.00302 (0.00161)
LASSO no Riesz	0.00890 (0.000409)	-0.00636 (0.000191)
LASSO ADML	0.00893 (0.000511)	-0.00585 (0.000830)
NNET no Riesz	0.00423 (0.000149)	-0.00317 (0.000208)
NNET ADML	**0.00443 (0.000287)**	**-0.00246 (0.000921)**
Panel B. Intensity load (log)		
Linear OLS	0.00622 (0.00123)	-0.00618 (0.00184)
Full OLS	0.0207 (0.00300)	-0.0250 (0.00343)
LASSO no Riesz	0.0183 (0.00120)	-0.0170 (0.000932)
LASSO ADML	0.0204 (0.00127)	-0.0153 (0.00167)
NNET no Riesz	0.0122 (0.000937)	-0.00673 (0.000501)
NNET ADML	**0.0152 (0.00104)**	**-0.00442 (0.00230)**

Note: Reported effects are cumulative average partial derivatives over lags 0-4; analytic state-clustered standard errors are in parentheses. Both panels use 5,643 first-differenced observations from 21 states, year and month-of-year fixed effects, precipitation controls, five state-level cross-fitting folds, and the same learner settings. Panel A uses equivalent days below THI 39 and at or above THI 69. Panel B uses the log intensity-duration measures defined in Section 3.3. For preferred NNET ADML, the Panel B Webb 95 percent intervals are [0.0131, 0.0174] for cold load and [-0.00927, 0.000157] for heat load.

Panel B of Table 4 gives the corresponding intensity-duration comparison. All six estimators produce a positive cold-load effect and a negative heat-load effect. Preferred NNET

ADML estimates are 0.0152 percentage points for `ln_THI39_Load` and -0.00442 percentage points for `ln_THI69_Load`. The cold-load Webb interval excludes zero, while the heat-load interval narrowly includes zero; the alternative measure therefore reinforces the directional composition pattern but provides weaker preferred-estimator evidence for the heat effect. Because fat content is in levels while the load measures are logs, the coefficients are semi-elastic responses rather than elasticities. Away from zero, the exact response to a 1 percent increase in raw load depends on its level; when load is large relative to one, it approaches a 0.0000442 percentage-point decline in fat content over the five-month response window. This magnitude is not directly comparable to the effect of an additional `THI69_Days` day.

Figure 4. Preferred NNET ADML cumulative effects on milk yield and fat content.



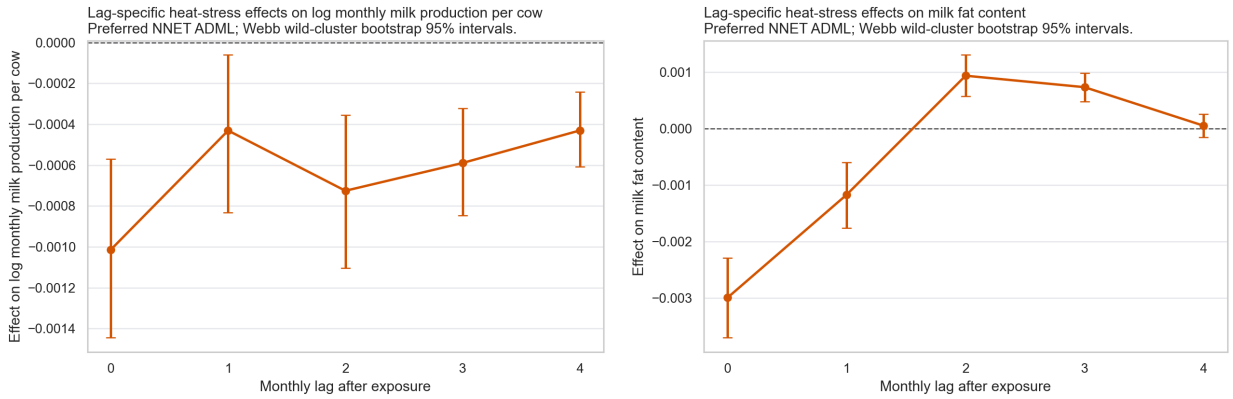
Note: Effects are cumulative over lags 0-4. Bars report point estimates, and whiskers report Webb wild-cluster-bootstrap 95 percent confidence intervals. Yield effects are percent changes because the outcome is logged; fat-content effects are percentage-point changes.

Figure 4 synthesizes our headline NNET ADML cumulative estimates across both production margins. In Panel A, one additional heat-stress day (`THI69_Days`) reduces monthly milk yield per cow by 0.32 percent (95 percent WCB interval [-0.48, -0.16] percent), an effect four times larger than the 0.08 percent reduction from a cold-stress day ([-0.14, -0.02] percent). In Panel B, heat stress lowers milk fat content by 0.00246 percentage points ([-0.00410, -0.00075]), whereas cold stress increases fat concentration by 0.00443 percentage points ([0.00389, 0.00502]). This comparison highlights a key physiological asymmetry: extreme heat depresses both output volume and milk quality, whereas extreme cold reduces output volume but concentrates butterfat solids.

6.3 Lag structure

Figure 5 separates the cumulative heat response into monthly lags. The quantity effect is largest in the month of exposure, but all five preferred lag estimates are negative and their Webb intervals exclude zero. This pattern indicates that heat damage is not confined to the contemporaneous month. The composition response is more front-loaded: fat content falls in months 0 and 1, partially rebounds in months 2 and 3, and is near zero at lag 4. The cumulative fat-content effect remains negative because the immediate decline is larger than the later rebound.

Figure 5. Lag-specific preferred NNET ADML heat-stress effects.



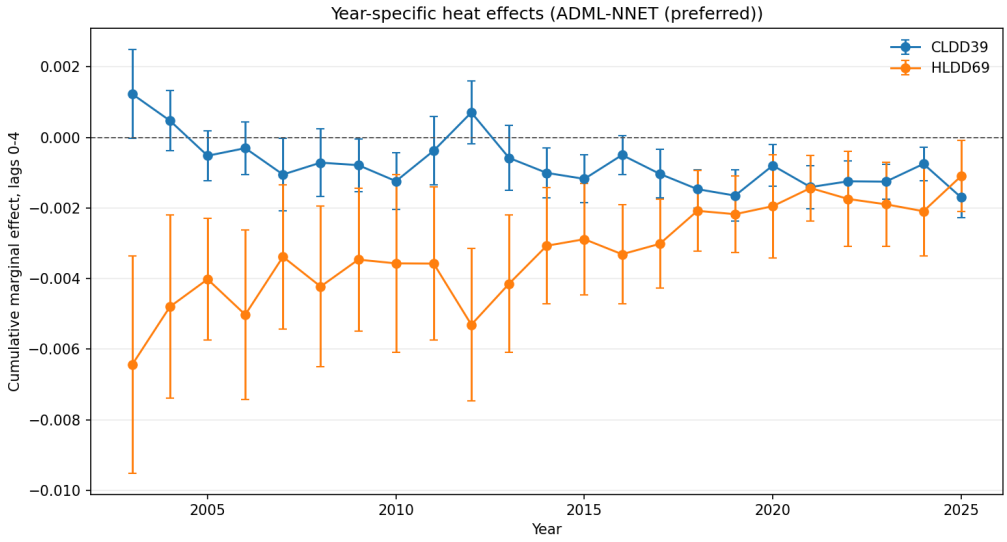
Note: Panel A reports log milk production per cow; Panel B reports fat content in percentage points. Whiskers are Webb wild-cluster-bootstrap 95 percent intervals. Given the Monte Carlo undercoverage, these intervals are interpreted as indicative.

7 Adaptation diagnostics

The main NNET ADML estimate for log production per cow is a pooled short-run heat effect. We use a complementary exercise to assess whether that response varies over time. The year-specific analysis averages preferred NNET ADML orthogonal scores by calendar year and remains tied to the main average-partial-derivative estimator.

Figure 6 reports year-specific cumulative NNET ADML results. The plotted coefficients are annual averages of the same debiased scores used to estimate the preferred pooled cumulative effects. The figure therefore answers a temporal-heterogeneity question: whether the marginal effect of heat stress appears stable over calendar time.

Figure 6. Year-specific cumulative marginal effects of cold and heat stress (2003-2025).



Note: Effects are summed over lags 0-4. The orange series is the policy-relevant heat-stress effect. The year-specific `THI69_Days` estimates are consistently negative early in the sample but become less negative in later years, especially after 2020. Confidence intervals show that year-specific estimates are noisier than the pooled estimate.

The year-specific exercise detects attenuation, but it cannot identify specific mechanisms such as cooling investment, genetics, herd composition, or management. If the attenuation reflects within-location adaptation, projecting the static 2003-2025 response forward would tend to overstate future damage. If it mainly reflects movement of production toward less vulnerable farms or regions, a projection that holds current production geography fixed may be closer to the relevant counterfactual. We favor the broader interpretation that technology and compositional change both contribute, but the evidence does not support imposing a single adaptation rate. We therefore do not mechanically extrapolate the attenuation estimates into the valuation.

8 Economic magnitude

8.1 Method and scope

We translate the preferred short-run productivity estimates into an annualized no-stress counterfactual sensitivity calculation for all 48 contiguous U.S. states. For heat, the counterfactual sets `THI69_Days` exposure to zero while holding cold exposure and other observed conditions fixed; for cold, it sets `THI39_Days` exposure to zero while holding heat exposure

and other conditions fixed. Because days in the THI 39-69 range form the omitted category, setting a stress measure to zero is interpreted as replacing that stress exposure with thermoneutral exposure rather than removing days from the calendar. The exercise expands the geographic accounting beyond the 21 states used to estimate the model and should be interpreted as a partial-equilibrium valuation of gross milk revenue, not as a separate causal model of prices, market equilibrium, or producer welfare.

The calculation makes four maintained assumptions:

1. Milk prices do not respond to the estimated production change.
2. Herd size and other production inputs remain fixed, and producers make no additional behavioral or technology adjustment within the valuation exercise.
3. The response estimated in 21 major dairy states is transportable to the other contiguous states.
4. The estimated average marginal response applies over the change from observed exposure to zero.

For state i and year y, let D_{iy} denote the sum of the 12 monthly `THI69_Days` or `THI39_Days` exposure measures, and let Q_{iy} denote observed USDA NASS annual milk production in pounds. We first define the annualized log effect of observed exposure as

$$\hat{s}_{iy} = \frac{\hat{\beta}_{cum}}{12} D_{iy},$$

where $\hat{\beta}_{cum}$ is the preferred cumulative lag-0-to-4 average partial derivative. Dividing by 12 maps the cumulative monthly response to average annual exposure. Predicted production under zero exposure to the selected stress measure is

$$\widehat{Q}^{0}_{iy} = Q_{iy} \exp\left(-\hat{s}_{iy}\right).$$

The physical shortfall is the difference between counterfactual no-stress and observed production:

$$\widehat{Q}^{loss}_{iy} = \widehat{Q}^{0}_{iy} - Q_{iy} = Q_{iy}\left[\exp\left(-\hat{s}_{iy}\right) - 1\right].$$

We value that shortfall as

$$\widehat{V}_{iy}^{loss} = \frac{\widehat{Q}_{iy}^{loss}}{100} P_{iy},$$

where P_{iy} is the corresponding state annual all-milk price received in dollars per hundredweight. This formulation is algebraically equivalent to the annualized loss-fraction calculation used to generate the reported results, so the estimates below are unchanged. Because $\hat{\beta}_{cum}$ is an average partial derivative from a flexible response model, applying it over the full movement from observed exposure to zero imposes a constant-average-marginal-response approximation.

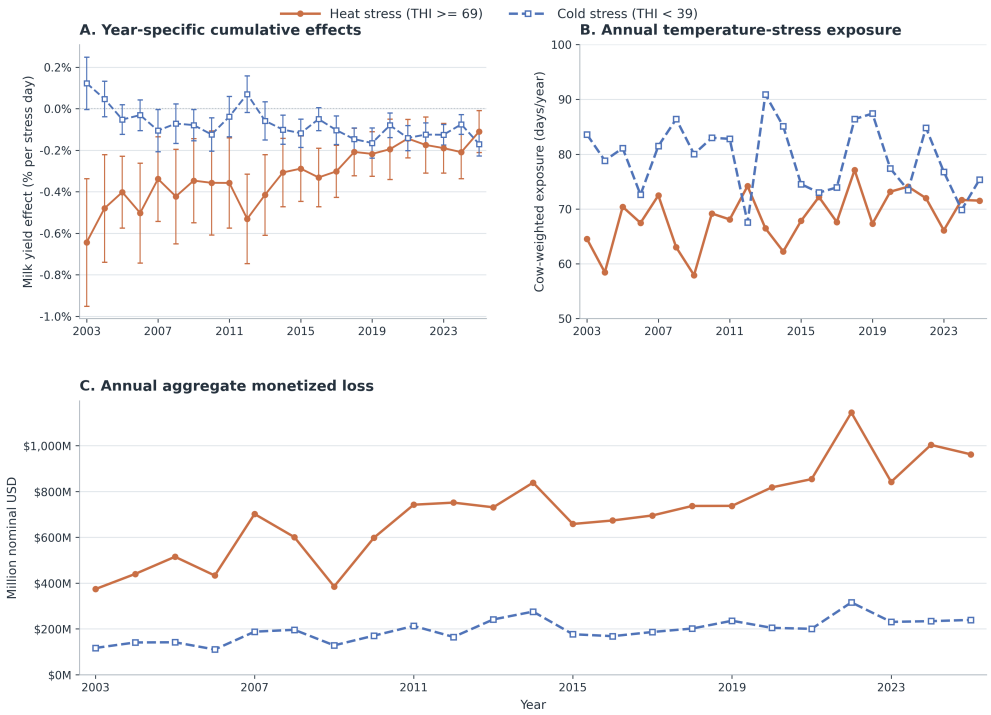
We repeat the full calculation using the endpoints of the preferred Webb wild-cluster-bootstrap interval to propagate coefficient sensitivity. The resulting bands do not incorporate uncertainty in weather exposure, production, prices, geographic transportability, the constant-response approximation, or general-equilibrium adjustments.

8.2 Volume losses

Across the contiguous United States, observed heat-stress exposure implies an estimated 87.744 billion pounds of lost milk during 2003-2025, valued at $16.233 billion in nominal dollars. The Webb coefficient-uncertainty range is $8.190 billion to $24.709 billion. Cold stress implies a further 24.084 billion pounds and $4.480 billion in nominal losses, with a Webb range of $0.941 billion to $8.220 billion. Heat therefore accounts for about four-fifths of the combined point-estimate value of temperature-stress losses in this accounting exercise.

Figure 7 connects three distinct parts of the economic accounting. Panel A reproduces the year-specific marginal effects from Figure 6 to place the valuation alongside the temporal-heterogeneity evidence. Panel B reports the cow-inventory-weighted annual frequency of heat- and cold-stress exposure. Panel C reports aggregate monetized losses using the pooled cumulative coefficients from the main model rather than the year-specific estimates in Panel A.

Figure 7. Annual temperature-stress effects, exposure, and aggregate economic losses (2003-2025).



Note: Panel A reports year-specific cumulative NNET ADML marginal effects for the 21-state estimation panel; whiskers are analytic 95 percent confidence intervals. Panel B reports cow-inventory-weighted 24-hour-equivalent annual heat- and cold-stress days on a focused 50-100 day scale. Panel C reports aggregate monetized losses based on the annualized pooled cumulative-coefficient valuation, exact NASS state annual production, and state annual all-milk prices. Panels B-C cover all 48 contiguous states for 2003-2024 and the 33 states with available annual production records in 2025. Panel A is a temporal-heterogeneity diagnostic and is not mechanically applied to the losses in Panel C. Monetary values are nominal.

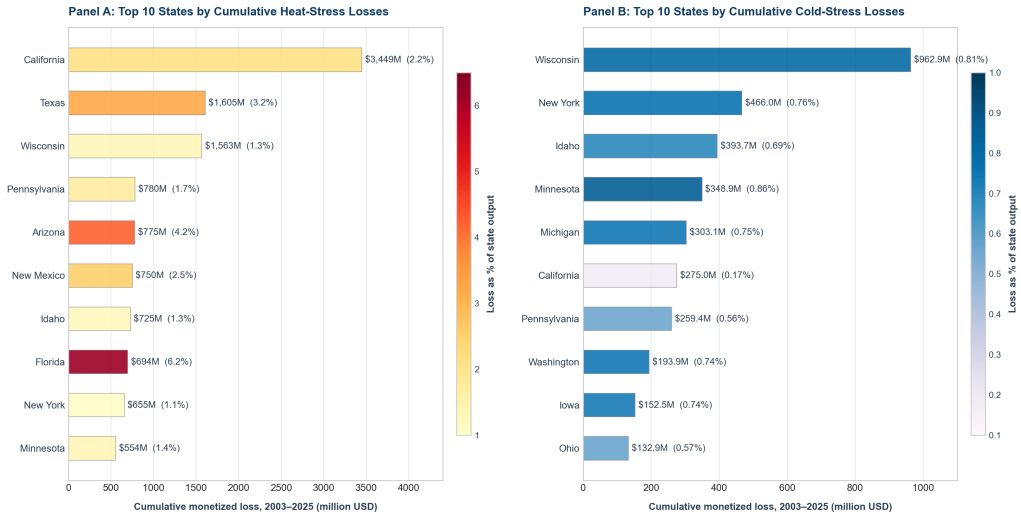
Panel B shows substantial year-to-year variation rather than a monotonic rise in temperature-stress exposure. Over the complete-coverage period 2003-2024, cow-weighted heat exposure averages 68.3 days per year and reaches its maximum of 77.1 days in 2018; cold exposure averages 79.6 days and peaks at 90.9 days in 2013. Despite the attenuation in the year-specific heat response shown in Panel A, aggregate losses in Panel C peak in 2022, at \$1.144 billion for heat and \$0.316 billion for cold stress, when high milk prices amplify the value of physical losses. Heat losses are next largest in 2024 (\$1.003 billion). The reported 2025 heat loss is \$0.962 billion, but it covers the 33 states with available annual production records and is not directly comparable to the complete 48-state totals through 2024. Figure 7 therefore should not be read as a pure temperature trend. Panel

C reflects exposure, milk-production scale, state composition, and nominal prices. These forces can raise monetized losses even when the marginal heat response in Panel A becomes less negative.

8.3 Spatial concentration

The geographic distribution reflects both exposure intensity and production scale. California has the largest cumulative heat loss ($3.449 billion), followed by Texas ($1.605 billion) and Wisconsin ($1.563 billion). Relative to state production, the heat burden is largest among the top-ranked states in Florida (6.17 percent), Arizona (4.24 percent), and Texas (3.25 percent). Cold losses are concentrated farther north: Wisconsin ($0.963 billion), New York ($0.466 billion), and Idaho ($0.394 billion) have the largest cumulative values.

Figure 8. States with the largest cumulative temperature-stress losses (2003-2025).



Note: Bars report million nominal USD. Parenthetical labels report physical lost milk as a percentage of total state milk production, and bar color encodes the same loss share. Panel A ranks heat-stress losses; Panel B ranks cold-stress losses.

Appendix Table A4 reports the national annual aggregate-loss series shown in Panel C of Figure 7. Appendix Tables A2 and A3 report the heat- and cold-state rankings underlying Figure 8.

9 Discussion and conclusion

9.1 What we learned

Temperature stress operates through two tails and two margins. One additional heat-stress day reduces milk production per cow by about 0.319 percent over five months and

lowers fat content by 0.00246 percentage points. Cold has a smaller negative quantity effect but a positive fat-content response. The agreement in direction across all six estimators makes this heat-cold asymmetry more robust than any single preferred magnitude. The lag profiles also show that volume losses persist beyond the exposure month, while the composition response is concentrated in the first two months.

The partial-equilibrium accounting places these effects at an economically meaningful scale: $16.233 billion in heat-related and $4.480 billion in cold-related gross milk-revenue losses across the contiguous United States during 2003-2025. Those totals are not welfare estimates, and the Webb ranges are sensitivity bands. They show where the order of magnitude comes from rather than supply exact confidence limits.

A core economic question is whether adaptation, combined with rising temperatures, ultimately reduces aggregate economic losses. The empirical evidence reveals a nuanced dynamic: while per-cow adaptation has attenuated the marginal yield penalty of an additional heat-stress day over time, the rising annual frequency of elevated THI days and expanding industry production scale prevent total economic losses from falling. Adaptation appears to function as a crucial defense that limits the growth of damages, but it does not seem to fully offset the compounding exposure of a warming climate and expanding dairy production scale.

9.2 Implications

For producers, loss shares identify where cooling and management investments may have the largest private return. Florida, Arizona, and Texas have the highest heat-loss shares among the leading states, while California has the largest absolute exposure because of its production scale. For cooperatives and processors, the simultaneous heat effects on volume and fat content imply exposure in both throughput and component value. For extension programs, the lag profile suggests that heat mitigation cannot be evaluated only against same-month production. At the national level, absolute losses and loss shares answer different questions: the former locate supply risk, while the latter identify regions where individual operations bear the greatest proportional burden.

9.3 Limitations

Four limitations bound the conclusions. First, state-level aggregation hides heterogeneity by herd size, breed, housing, and cooling technology. Second, the valuation fixes prices and omits equilibrium responses, producer surplus, and a direct fat-component revenue calculation. Third, the 21-cluster Monte Carlo shows that preferred-estimator intervals undercover, even after the Riesz correction. Fourth, the adaptation diagnostics cannot distinguish technology adoption from genetics, herd composition, management, or spatial reallocation. These limitations make the sign and approximate magnitude more credible than precise interval endpoints or mechanism claims.

9.4 Extensions

Farm- or herd-level microdata would permit direct tests of management and herd-size heterogeneity. Protein, somatic-cell, fertility, mortality, and culling outcomes would broaden the composition and animal-health margins. A component-pricing module could translate fat and protein responses into revenue separately from milk volume. Climate-projection exercises should condition damages on explicit adaptation scenarios rather than apply a fixed historical coefficient indefinitely.

Overall, the evidence shows that thermal stress is not only a heat-and-volume problem. Heat damages quantity and composition together; cold reduces quantity but shifts composition in the opposite direction. Flexible orthogonal estimation recovers nonlinear response magnitudes, cow-weighted spatial aggregation specifies how weather exposure aligns with herd location, and the validation exercise demonstrates why empirical uncertainty requires cautious interpretation.

Appendix

Appendix A. Estimation details

Basis construction and standardization

We construct the coarse THI-bin basis and short-run basis expansion in levels before differencing. The basis expansion includes current-month through four-month-lagged THI exposure, matching precipitation lags, third-degree polynomial terms, within-lag THI-by-precipitation interactions, and time controls. Levels basis functions are standardized with

training-fold moments. The mean cancels in first differences, and analytical derivatives are scaled by the same training-fold standard deviations (S. Klosin and M. Vilgalys, 2025).

First-stage learners

Let Δb_{it} denote the standardized differenced basis vector. The first-stage regression estimates

$$\Delta\hat{\gamma}_{it} = \Delta b_{it}'\hat{\beta}. \tag{A1}$$

The official LASSO implementation penalizes the complete 88-column basis expansion, including 55 weather terms and 33 time-control indicators, with regularization selected by state-respecting cross-validation. LASSO ADML adds the Riesz correction; LASSO no Riesz reports the corresponding plug-in derivative. The preferred flexible learner is a ReLU neural network. NNET ADML combines the cross-fitted outcome learner with the Riesz correction, while NNET no Riesz isolates the plug-in result. Experimental unpenalized-time-effect LASSO and smooth-activation neural-network variants were evaluated during model development but are not promoted in the six-estimator main comparison.

Analytical derivatives and Riesz correction

Let $b_{D,\ell}(D_{it}, X_{it})$ denote the analytical derivative of the basis vector with respect to exposure at lag ℓ. We compute derivatives analytically,

$$\frac{\partial\Delta\hat{\gamma}_{it}}{\partial D_{i,t-\ell}} = b_{D,\ell}(D_{it}, X_{it})'\hat{\beta}, \tag{A2}$$

rather than with numerical perturbations. We estimate a sparse linear approximation to the Riesz representer with the convex minimum-distance problem in S. Klosin and M. Vilgalys (2025). If $\hat{\alpha}_{it}$ is the estimated representer, the held-out orthogonal score is

$$\hat{\psi}_{\ell,it} = \frac{\partial\Delta\hat{\gamma}_{it}}{\partial D_{i,t-\ell}} + \hat{\alpha}_{it}\left(\Delta y_{it} - \Delta\hat{\gamma}_{it}\right), \tag{A3}$$

and the estimator averages these scores over states and months,

$$\hat{\tau}_{\ell} = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\hat{\psi}_{\ell,it}. \tag{A4}$$

Orthogonality reduces sensitivity to small nuisance-estimation errors when the relevant product-rate conditions hold (V. Chernozhukov et al., 2022; S. Klosin and M. Vilgalys, 2025).

Inference

Analytic variance estimates allow arbitrary serial correlation within states. Full OLS derivative scores include both observation-level derivative variation and the coefficient influence-function contribution implied by the OLS normal equations. The Webb wild-cluster bootstrap applies six-point state-level weights to the already-computed ADML or corrected OLS scores without refitting nuisance functions (A. C. Cameron et al., 2008; M. D. Webb, 2023). The Monte Carlo evidence in Section 5 shows why these intervals are treated as indicative in the empirical application.

Appendix B. Main-estimate robustness

Table A1 reports robustness checks for the preferred NNET ADML estimator. The first row repeats the main lag 0-4 specification. The next two rows vary the distributed-lag window while keeping the baseline year and calendar-month time controls. The final row keeps the promoted lag 0-4 window but replaces the separate year and month-of-year fixed effects with full month-year fixed effects, so identification comes from cross-state weather variation within the same calendar month. We also report a specification that adds wind speed and surface solar radiation as additional lagged polynomial weather controls.

Table A1. Preferred NNET ADML robustness checks.

Specification	Lag	Time controls	Added weather controls	N	THI39_Days effect (SE)	THI69_Days effect (SE)
Main specification	0-4	Year and month-of-year fixed effects	None	5,691	-0.000782 (0.000293)	-0.00319 (0.000773)
Shorter lag window	0-3	Year and month-of-year fixed effects	None	5,712	-0.000814 (0.000267)	-0.00284 (0.000737)
Longer lag window	0-5	Year and month-of-year fixed effects	None	5,670	-0.00103 (0.000270)	-0.00319 (0.000699)
Month-year fixed effects	0-4	Full month-year fixed effects	None	5,691	-0.00114 (0.000297)	-0.00327 (0.000749)
Additional weather controls	0-4	Year and month-of-year fixed effects	Wind speed and solar radiation	5,691	-0.000732 (0.000252)	-0.00361 (0.000729)

Note: The reported effects are cumulative average partial derivatives for `THI39_Days` and `THI69_Days`. Parentheses report analytic state-clustered standard errors. The corresponding Webb wild-cluster-bootstrap intervals are also negative for `THI69_Days` in all four specifications. `THI39_Days` represents cold-stress exposure days below THI 39; `THI69_Days` represents heat-stress exposure days at or above THI 69.

The preferred heat estimate is stable across these checks. Shortening the lag window from 0-4 to 0-3 makes the cumulative `THI69_Days` effect slightly less negative, while extending it to 0-5 leaves the estimate almost unchanged. Full month-year fixed effects also leave the heat effect close to the main estimate. Adding wind speed and solar radiation makes the preferred heat effect modestly more negative but does not change the qualitative conclusion. These results support the promoted lag 0-4 specification and show that the headline heat-stress estimate is not driven by the baseline time-control choice or by omission of these additional weather controls.

Appendix C. Economic loss valuation details

Tables A2-A4 report the state and national components of the annualized valuation described in the empirical-results section. The heat and cold calculations use the preferred NNET ADML cumulative coefficients of -0.003189 and -0.000782, respectively, and exact NASS annual production and state annual prices. State rankings are based on the point-estimate monetized loss.

Table A2. State-level cumulative heat-stress economic losses, ranked by monetized loss (top 20 states, 2003-2025).

State	Lost milk	Loss share	Value	Loss/cow	Value/cow/yr
California	20.000	2.16%	3,449.0	1,959	85
Texas	8.327	3.25%	1,604.8	3,542	154
Wisconsin	8.411	1.31%	1,563.4	1,242	54
Pennsylvania	4.015	1.68%	779.8	1,507	66
Arizona	4.366	4.24%	775.3	4,988	217
New Mexico	4.398	2.54%	749.7	2,466	107
Idaho	3.985	1.26%	725.2	1,227	53
Florida	3.180	6.17%	693.5	7,158	311
New York	3.409	1.07%	654.7	1,049	46
Minnesota	2.946	1.36%	553.5	1,291	56
Michigan	2.910	1.31%	531.8	1,364	59
Ohio	2.239	1.83%	428.4	1,658	72
Iowa	1.920	1.72%	358.1	1,901	83
Kansas	1.867	2.55%	340.4	3,637	158
Indiana	1.766	2.00%	333.6	2,356	102
Georgia	1.600	4.21%	328.4	6,356	276
Colorado	1.112	1.28%	211.6	1,473	64
Washington	1.086	0.77%	204.7	801	35
Illinois	1.007	2.37%	185.9	2,189	95
South Dakota	0.915	1.53%	182.4	1,894	82

Note: Lost milk is in billion pounds; value is in million nominal USD; loss per cow is cumulative USD per average annual cow inventory; value per cow per year is annualized USD. Loss share is lost milk divided by total state milk production. Calculations use the annualized cumulative `THI69_Days` coefficient and exact NASS annual production and state annual prices.

Table A3. State-level cumulative cold-stress economic losses, ranked by monetized loss (top 10 states, 2003-2025).

State	Lost milk	Loss share	Value
Wisconsin	5.169	0.81%	962.9
New York	2.438	0.76%	466.0
Idaho	2.185	0.69%	393.7
Minnesota	1.863	0.86%	348.9
Michigan	1.655	0.75%	303.1
California	1.618	0.17%	275.0
Pennsylvania	1.339	0.56%	259.4
Washington	1.051	0.74%	193.9
Iowa	0.819	0.74%	152.5
Ohio	0.697	0.57%	132.9

Note: Lost milk is in billion pounds and value is in million nominal USD. Loss share is lost milk divided by total state milk production. Calculations use the annualized cumulative `THI39_Days` coefficient and exact NASS annual production and state annual prices.

Table A4. National annual contiguous U.S. heat- and cold-stress economic losses (2003-2025).

Year	Heat lost milk	Heat value	Cold lost milk	Cold value
2003	2.984	373.8	0.917	116.9
2004	2.736	440.0	0.861	140.8
2005	3.365	514.8	0.919	141.6
2006	3.333	433.0	0.844	110.2
2007	3.636	701.6	0.970	187.9
2008	3.253	600.7	1.049	196.1
2009	2.984	385.1	0.978	128.2
2010	3.607	597.8	1.032	170.6
2011	3.641	742.1	1.043	212.8
2012	4.047	751.2	0.871	164.9
2013	3.636	730.8	1.184	241.3
2014	3.503	838.6	1.124	275.5
2015	3.821	658.0	1.011	176.7
2016	4.114	673.3	1.019	168.3
2017	3.929	695.2	1.046	186.5
2018	4.512	736.7	1.234	201.7
2019	3.942	737.0	1.258	235.5
2020	4.513	817.8	1.113	204.6
2021	4.611	853.8	1.071	200.3
2022	4.498	1,144.2	1.244	315.8
2023	4.121	841.9	1.137	230.7
2024	4.431	1,003.2	1.034	233.8
2025	4.526	962.1	1.126	239.3

Note: Lost milk columns are in billion pounds and value columns are in million nominal USD. The full-period totals are 87.744 billion pounds and $16.233 billion for heat stress, and 24.084 billion pounds and $4.480 billion for cold stress. Webb coefficient-uncertainty ranges are reported in the empirical-results section.

References

F. Annan and W. Schlenker. (2015). Federal crop insurance and the disincentive to adapt to extreme heat. *American Economic Review*.

D. V. Armstrong. (1994). Heat stress interaction with shade and cooling. *Journal of Dairy Science*.

M. Burke and K. Emerick. (2016). Adaptation to climate change: Evidence from U.S. agriculture. *American Economic Journal: Economic Policy*.

B. Callaway, A. Goodman-Bacon, and P. H. C. Sant'Anna. (2024). Event studies with a continuous treatment. *AEA Papers and Proceedings*.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*.

V. Chernozhukov, W. K. Newey, and R. Singh. (2022). Automatic debiased machine learning

of causal and structural effects. *Econometrica*.

E. Choi, V. C. Souza, J. A. Dillon, E. Kebreab, and N. D. Mueller. (2024). Comparative analysis of thermal indices for modeling cold and heat stress in US dairy systems. *Journal of Dairy Science*.

O. Deschênes and M. Greenstone. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*.

X. Fan and J. Ma. (2024). A spatial econometric analysis of weather effects on milk production. *Earth*.

A. Garcia and K. Kalscheur. (2004). Dairies and local economic development. South Dakota State University Cooperative Extension Service.

M. Gisbert-Queral, A. Henningsen, B. Markussen, M. T. Niles, E. Kebreab, S. Rabotyagov, and N. D. Mueller. (2021). Climate impacts and adaptation in US dairy systems 1981–2018. *Nature Food*.

H. Górska-Warsewicz, K. Rejman, W. Laskowski, and M. Czeczotko. (2019). Milk and dairy products and their nutritional contribution to the average polish diet. *Nutrients*.

R. P. Heaney. (2014). How sound is the science behind the dietary recommendations for dairy? *The American Journal of Clinical Nutrition*.

J. M. L. Heck, H. J. F. van Valenberg, J. Dijkstra, and A. C. M. van Hooijdonk. (2009). Seasonal variation in the dutch bovine raw milk composition. *Journal of Dairy Science*.

S. Hsiang. (2016). Climate econometrics. *Annual Review of Resource Economics*.

J. Hutchins, M. Skidmore, and D. Nolan. (2025). Vulnerability of US dairy farms to extreme heat. *Food Policy*.

International Dairy Foods Association (IDFA). (2025). Dairy delivers: 2025 economic impact study of the dairy products industry.

C. T. Kadzere, M. R. Murphy, N. Silanikove, and E. Maltz. (2002). Heat stress in lactating dairy cows: A review. *Livestock Production Science*.

N. Key, S. Sneeringer, and D. Marquardt. (2014). Climate change, heat stress, and u.s. Dairy production. USDA Economic Research Service.

S. Klosin and M. Vilgalys. (2025). Estimating continuous treatment effects in panel data using machine learning with a climate application.

R. A. Lopez, C. Laughton, D. Kim, and H. S. Suh. (2022). Climate change and productivity of northeastern dairy farms. *Agricultural and Resource Economics Review*.

G. Mauger, Y. Bauman, T. Nennich, and E. Salathé. (2014). Impacts of climate change on milk production in the united states. *The Professional Geographer*.

J. M. Mbuthia, A. Eggert, and N. Reinsch. (2022). Cooling temperature humidity index-days as a heat load indicator for milk production traits. *Frontiers in Animal Science*.

T. M. McWhorter, M. Sargolzaei, C. G. Sattler, M. D. Utt, S. Tsuruta, I. Misztal, and D. Lourenco. (2023). Single-step genomic predictions for heat tolerance of production yields in US holsteins and jerseys. *Journal of Dairy Science*.

S. S. Moore, A. Costa, M. Penasa, S. Callegaro, and M. De Marchi. (2023). How heat stress conditions affect milk yield, composition, and price in italian holstein herds. *Journal of Dairy Science*.

D. Mukherjee, B. E. Bravo-Ureta, and A. De Vries. (2012). Dairy productivity and climatic conditions: Econometric evidence from south-eastern united states. *Australian Journal*

of Agricultural and Resource Economics.

OECD/FAO. (2022). OECD-FAO agricultural outlook 2022–2031. OECD Publishing, Paris/Food; Agriculture Organization of the United Nations, Rome.

J. Prieto, Z. Lin, K. F. Reed, C. A. Wolf, and A. Ortiz-Bobea. (2026). Milk composition responses amplify economic damages from heat stress. *Environmental Research Letters*.

O. Ravagnolo, I. Misztal, and G. Hoogenboom. (2000). Genetic component of heat stress in dairy cattle, development of heat index function. *Journal of Dairy Science*.

I. J. Salfer, C. D. Dechow, and K. J. Harvatine. (2019). Annual rhythms of milk and milk fat and protein production in dairy cattle in the united states. *Journal of Dairy Science*.

W. Schlenker and M. J. Roberts. (2009). Nonlinear temperature effects indicate severe damages to u.s. Crop yields under climate change. *Proceedings of the National Academy of Sciences*.

N. R. St-Pierre, B. Cobanov, and G. Schnitkey. (2003). Economic Losses from Heat Stress by US Livestock Industries. *Journal of Dairy Science*.

J. Tack, A. Barkley, and L. L. Nalley. (2015). Effect of warming temperatures on US wheat yields. *Proceedings of the National Academy of Sciences*.

USDA Economic Research Service. (2023). Dairy data: Per capita consumption of selected dairy products. https://www.ers.usda.gov/data-products/dairy-data/.

USDA National Agricultural Statistics Service (NASS). QuickStats database.

W. Vroege, T. Dalhaus, E. Wauters, and R. Finger. (2023). Effects of extreme heat on milk quantity and quality. *Agricultural Systems*.

M. D. Webb. (2023). Reworking wild bootstrap-based inference for clustered errors. *Canadian Journal of Economics/Revue canadienne d'économique*.

J. W. West. (2003). Effects of heat-stress on production in dairy cattle. *Journal of Dairy Science*.

E. Zaveri and D. B. Lobell. (2019). The role of irrigation in changing wheat yields and heat sensitivity in india. *Nature Communications*.