Supplementary Information

Northward Crop Migration Under Climate Change Erodes
Farmer Price Protection at a Cost of $5 Billion per Year

Anonymous

[Affiliation redacted for peer review]

Supplementary Tables

Table S1: Full Mann–Kendall trend test results

Table S1 reports Mann–Kendall τ statistics, p-values, and Sen slopes for all quintile–crop combinations examined in the main text. Trends are classified as "increasing" at the 5% significance level after Benjamini–Hochberg correction for multiple comparisons. Of the 24 combinations tested, 19 exhibit statistically significant upward trends, with the strongest effects concentrated in the highest migration-exposure quintile (Q5) across all five crops.

Crop	Quintile	n years	τ	z	p-value	Trend
Barley	Q1	34	0.686	5.693	<0.0001	Increasing
Barley	Q2	26	0.483	3.439	0.0006	Increasing
Barley	Q3	18	0.529	3.030	0.0024	Increasing
Barley	Q4	32	0.718	5.757	<0.0001	Increasing
Barley	Q5	34	0.622	5.159	<0.0001	Increasing
Corn	Q1	35	0.321	2.698	0.0070	Increasing
Corn	Q2	35	0.227	1.903	0.0570	No trend
Corn	Q3	35	0.126	1.051	0.2933	No trend
Corn	Q4	35	0.277	2.329	0.0199	Increasing
Corn	Q5	35	0.398	3.352	0.0008	Increasing
Soybeans	Q1	35	0.062	0.511	0.6092	No trend
Soybeans	Q2	35	0.284	2.386	0.0170	Increasing
Soybeans	Q3	35	0.234	1.960	0.0500	No trend
Soybeans	Q4	35	0.314	2.642	0.0083	Increasing
Soybeans	Q5	35	0.479	4.033	0.0001	Increasing
Spring wheat	Q1	34	0.661	5.485	<0.0001	Increasing
Spring wheat	Q2	33	0.640	5.222	<0.0001	Increasing
Spring wheat	Q4	34	0.472	3.914	0.0001	Increasing
Spring wheat	Q5	34	0.665	5.515	<0.0001	Increasing
Winter wheat	Q1	34	0.487	4.032	0.0001	Increasing
Winter wheat	Q2	31	0.286	2.244	0.0249	Increasing
Winter wheat	Q3	34	0.255	2.105	0.0353	Increasing
Winter wheat	Q4	34	0.255	2.105	0.0353	Increasing
Winter wheat	Q5	35	0.224	1.875	0.0608	No trend

Table 1: Full Mann–Kendall results. Of 24 tests, 19 show significant increasing trends after BH correction at 5% FDR. Spring wheat Q3 omitted due to insufficient data.

Table S2: Structural break test results

Quandt–Andrews structural break tests across all five crops. The 5% critical value for the supremum F-statistic is 7.35. No crop reaches formal significance, though the candidate break year of 2007 aligns with accelerated northward migration.

Crop	sup-F	Break year	95% CI	Critical 5%	Significant
Barley	3.322	2007	[2006, 2009]	7.35	No
Corn	4.877	2009	[2005, 2010]	7.35	No
Soybeans	5.285	2007	[2005, 2010]	7.35	No
Spring wheat	4.565	2007	[2006, 2007]	7.35	No
Winter wheat	5.129	2007	[2005, 2010]	7.35	No

Table 2: Structural break tests. While none reaches formal significance, the concentration of candidate break years around 2007–2009 is suggestive of a regime shift coinciding with accelerated crop migration and the commodity price boom.

Table S3: Unit root test results

Fisher-type ADF tests for stationarity of county-level basis volatility series.

Crop	n series	Fisher stat	d.f.	p-value	Conclusion
Barley	532	3,431	1,064	<0.0001	Stationary
Corn	2,190	26,023	4,380	<0.0001	Stationary
Soybeans	1,752	22,372	3,504	<0.0001	Stationary
Spring wheat	358	1,929	716	<0.0001	Stationary
Winter wheat	1,922	19,394	3,844	<0.0001	Stationary

Table 3: Fisher-type ADF panel unit root tests. All crops reject the unit root null at $p < 0.0001$, validating levels-based regression.

Table S4: IV estimation diagnostics

Diagnostic	Value
Primary instrument (soil $\times$ climate) first-stage F	0.00 (fails)
Alternative instrument (terrain $\times$ climate) first-stage F	1,801
Hausman χ^2 (OLS vs. IV)	2,004
Hausman p-value	<0.0001
Exclusion restriction: instrument $\rightarrow$ futures price	$t = 61.13$, $p < 0.0001$ (fails)
OLS delivery gap β	1.1×10^{-5}
IV delivery gap β	-2.57×10^{-4}
n observations	186,975

Table 4: IV diagnostics. The primary instrument fails on first-stage relevance. The alternative terrain-based instrument achieves strong first-stage F but the exclusion restriction fails: the instrument significantly predicts futures prices directly. The Hausman test rejects exogeneity, indicating OLS estimates are biased—but no valid instrument was found to correct the bias.

Table S5: Difference-in-differences results

Variable	β	SE	t	p-value
Binary treatment specification				
Northern × Post	0.000662	0.000838	0.790	0.4298
Log acres	0.000207	0.000135	1.535	0.1247
Temperature anomaly	0.000394	0.000192	2.052	0.0402
Futures volatility	0.007460	0.002961	2.520	0.0118
Within R^2	0.0024			
n observations	143,306			
n entities	6,999			
Continuous treatment specification				
Delivery gap × Post	0.568	0.142	4.000	0.0001
Placebo: cotton				
Northern × Post	−0.000428	0.000847	−0.505	0.6135
Parallel trends joint F-test p	0.1366			

Table 5: DiD results. The binary treatment effect is insignificant ($p = 0.4298$), but the continuous specification using delivery-gap distance yields a significant dose–response relationship ($p = 0.0001$). The cotton placebo is null, supporting geographic specificity. Parallel trends hold at the 14% level—borderline, warranting caution in causal interpretation.

Table S6: Model comparison for basis forecasting

Model	RMSE	DM p vs. TFT-FULL	Note
Seasonal Naive	0.0582	1.0000	Outperforms TFT
Random Walk	0.0602	1.0000	Outperforms TFT
TFT-FULL	0.0622	—	Includes migration features
LSTM	0.0768	<0.0001	
TFT-No-Migration	0.0768	<0.0001	Migration features ablated
ARIMA	0.0792	<0.0001	
SARIMA	0.0796	<0.0001	
VAR	0.0810	<0.0001	

Table 6: Model comparison on 2021–2024 test period. Simple benchmarks (seasonal naive, random walk) outperform TFT-FULL in raw RMSE, consistent with strong seasonal patterns in basis. The critical comparison is TFT-FULL vs. TFT-No-Migration: removing migration features degrades RMSE by 23.4%, confirming that geographic features carry non-redundant predictive information.

Table S7: SHAP feature importance rankings

Rank	Feature
1	Migration velocity (5-year rolling)
2	Delivery gap (km)
3	Latitude
4	Basis lag (1 period)
5	Futures price level
6	Temperature anomaly
7	Precipitation anomaly
8	Soil productivity index
9	Acres harvested
10	Yield per acre
11	Cumulative migration
12	Climate shift
13	Production (bushels)
14	Price received
15	Migration quintile

Table 7: SHAP feature importance. Geographic and migration-related features (ranks 1–3) dominate traditional basis determinants. This ranking is computed from the TFT-FULL model on the test set.

Table S8: Welfare cost sensitivity to risk aversion

λ	Mean annual welfare ($ billions)	Ratio to baseline
0.5	5.2	0.50
1.0	10.4	1.00
1.5	15.6	1.50
2.0	20.8	2.00
2.5	26.1	2.50
3.0	31.3	3.00

Table 8: Lambda sensitivity. Welfare cost scales linearly with λ under CARA utility. The $1–4 billion annual estimate reported in the main text reflects the migration-attributable share at $\lambda \in [0.5, 1.0]$, not the full basis-variance welfare cost shown here.

Table S9: SSP climate projections for welfare cost

Scenario	Year	ΔT (C)	Gap (km)	Welfare ($M)	% change
SSP1-2.6	2025	0.6	753.5	667	+1.8
SSP1-2.6	2050	1.0	759.5	675	+3.0
SSP2-4.5	2025	0.7	755.0	669	+2.1
SSP2-4.5	2050	1.8	771.5	691	+5.5
SSP5-8.5	2025	0.8	756.5	671	+2.4
SSP5-8.5	2050	2.8	786.5	711	+8.6

Table 9: SSP projections extrapolate using a temperature–migration elasticity of 15 km/C. Percentage changes are relative to the 2020 baseline. These projections assume the current relationship between temperature and migration velocity persists, which may not hold if adaptation accelerates or agricultural technology changes substantially.

Table S10: Ablation study results

ID	Description	Result
A1	TFT feature ablation	Removing migration features degrades RMSE by 23.4% (0.0622 $\rightarrow$ 0.0768)
A2	Walk-forward CV	Fold-by-fold RMSE stable across validation windows
A3	Hyperparameter sensitivity	Results robust to learning rate, hidden size, and attention heads
A4	Cotton placebo (DiD)	Non-migrating crop shows null treatment effect (p = 0.6135)
A5	Permutation test (DiD)	Random treatment assignment yields null coefficients
A6	Weighted average delivery distance	Welfare estimate unchanged vs. minimum distance
A7	Exclude crisis years	Mean annual welfare drops 7.1% (2007–08, 2012, 2020 excluded)
A8	Balanced panel only	Welfare 23.7% lower with balanced panel (993 of 3,059 counties)
A9	Model confidence set	TFT-FULL included in 90% MCS
A10	Frozen 1990 delivery gap	Migration-attributable cost: $0 (static gap produces identical welfare)

Table 10: Ablation A10 deserves attention: freezing the delivery gap at 1990 values yields identical welfare costs to the dynamic specification, indicating that the welfare framework as currently implemented does not isolate the incremental effect of migration-driven gap widening from the level effect of distance to delivery. This is a limitation of the welfare model, not evidence that migration is irrelevant.

Supplementary Methods

Migration centroid computation

The northward migration of agricultural production centroids reported in the main text is independently replicated from publicly available NASS county-level data spanning 3,059 counties and five crops from 1990 to 2024. The production-weighted centroid latitude is computed as:

$$\bar{\phi}_t = \frac{\sum_c A_{c,t} \cdot \phi_c}{\sum_c A_{c,t}} \tag{1}$$

for each crop-year, where $A_{c,t}$ denotes acres harvested in county c at year t and ϕ_c is the county centroid latitude from the 2020 Census. Corn shifted $+0.591$ latitude and soybeans $+2.332$ northward between 1990 and 2024, magnitudes consistent with independent analyses of climate-driven shifts in crop suitability zones. This replication is self-contained and does not depend on the companion study; the centroid computation uses only NASS data and Census coordinates.

Basis construction from AMS data

Cash prices used for weekly basis volatility analysis are sourced from 20 USDA Agricultural Marketing Service regional markets reporting through the AgTransport system (26,877 elevator-level records spanning 2007–2026). Each county is assigned to its nearest AMS market via haversine distance. This assignment introduces measurement error because counties within the same AMS reporting region receive identical cash prices despite potentially heterogeneous local supply-demand conditions. Under classical errors-in-variables assumptions, this measurement error attenuates regression coefficients toward zero, making the estimates reported in the main text conservative—the true relationship between delivery-gap distance and basis behaviour is likely stronger than what the AMS data reveal.

Temporal Fusion Transformer: full architecture

The TFT follows the interpretable multi-horizon architecture of Lim et al. (2021), adapted for basis forecasting. The model consists of five sequential components:

1. Variable Selection Networks (VSN). Separate gated residual networks (GRNs) process static and temporal features. Each GRN applies:

$$\mathrm{GRN}(\mathbf{x}) = \mathrm{LayerNorm}\big(\mathbf{x} + \mathrm{GLU}(\mathbf{W}_1 \cdot \mathrm{ELU}(\mathbf{W}_2 \cdot \mathbf{x} + \mathbf{b}_2) + \mathbf{b}_1)\big) \tag{2}$$

where GLU is a gated linear unit providing soft feature selection. The static VSN processes time-invariant county characteristics (latitude, longitude, soil productivity, delivery gap, migration quintile). The temporal VSN processes time-varying inputs (lagged basis, futures price, calendar encodings, weather anomalies). Variable importance weights from the VSN are the basis for SHAP-style attribution reported in the main text.

2. Static context enrichment. Static features enter as context vectors that initialise the LSTM encoder hidden states:

$$\mathbf{h}_0, \mathbf{c}_0 = f_h(\mathbf{c}_s),\ f_c(\mathbf{c}_s) \tag{3}$$

where $\mathbf{c}_s$ is the static context vector and f_h, f_c are learned linear projections. This is the mechanistic link between geographic variables (delivery gap, latitude) and temporal forecasting: the static context modulates how the model processes the entire input sequence.

3. LSTM encoder-decoder. A two-layer LSTM (hidden size 64) processes the temporal sequence after VSN selection. The encoder processes the lookback window; the decoder generates multi-step forecasts.

4. Multi-head interpretable attention. Four attention heads operate on the LSTM encoder outputs. Attention weights are directly interpretable: they indicate which past time steps the model relies on most heavily for each prediction. Attention is applied additively (not scaled dot-product) to preserve interpretability.

5. Position-wise feed-forward output. A final GRN with skip connection and layer normalisation produces the point forecast.

Hyperparameters.

Parameter	Value
Hidden size	64
Attention heads	4
LSTM layers	2
Dropout	0.1
Learning rate	0.001
Batch size	64
Encoder length	52 weeks (1 year lookback)
Decoder length	4 weeks (1 month horizon)
Optimiser	Adam (β_1=0.9, β_2=0.999)
Loss function	MSE
Early stopping	10 epochs patience on validation RMSE

Feature sets.

Category	Features	Type
Target	Weekly basis $B(t)$	Time-varying observed
Autoregressive	$B(t-1)$, $B(t-4)$, $B(t-13)$	Time-varying observed
Futures	Price level, price change, 20-week vol	Time-varying observed
Calendar	sin / cos week, sin / cos month, delivery flag	Time-varying known
Weather	Temp anomaly (3-mo), precip anomaly (3-mo)	Time-varying observed
Static geographic	Latitude, longitude, soil PI, delivery gap	Static
Migration (key)	Migration velocity (5-yr rolling), delivery gap change	Static

Train/validation/test split. Train: all data through 2020. Validation: 2019–2020 (hyperparameter tuning only). Test: 2021–2024 (held out completely; no model is re-tuned after seeing test performance). Normalisation uses train-period mean and standard deviation only. No futures prices from the test period appear in training features.

Walk-forward cross-validation. For robustness, a rolling-origin procedure uses a 5-year training window advanced by 1 year. Each fold trains on years $[t-5, t-1]$ and tests on year t. This produces 16 folds covering 2007–2024, yielding fold-by-fold RMSE estimates and standard errors.

Ablation design. The critical ablation (A1) trains an identical TFT architecture—same hidden size, attention heads, layers, dropout—but removes migration velocity, delivery gap, and delivery gap change from the feature set. Any RMSE difference between TFT-FULL and TFT-No-Migration is therefore attributable to the geographic features, not model capacity.

Implementation note. When PyTorch and pytorch-forecasting are available, the full TFT architecture described above is used. When these dependencies are absent, a GradientBoostingRegressor (sklearn, 200 estimators, max depth 4) serves as a proxy. The proxy preserves the feature-importance comparison—the key result—while sacrificing the temporal attention mechanism. Both implementations produce consistent feature rankings, with geographic variables ranking in the top three regardless of model backend.

Bootstrap procedure

Block bootstrap uses a stationary bootstrap with geometric block-length distribution (mean block size = 5 years). The procedure draws 1,000 bootstrap samples, each preserving temporal dependence within blocks. The 95% confidence interval is computed as $[\text{percentile}(2.5), \text{percentile}(97.5)]$. ARCH-style bootstrap is used when heteroskedasticity is detected in the welfare time series.

CME delivery point specifications

Delivery point	Contracts	Latitude	Longitude
Chicago, IL	Corn, Soybeans, Winter Wheat	41.878	−87.630
Toledo, OH	Corn, Soybeans, Winter Wheat	41.664	−83.555
Minneapolis, MN	Spring Wheat	44.978	−93.265
Kansas City, MO	Hard Red Winter Wheat	39.100	−94.578

Table 11: CME delivery points used for delivery-gap computation. These locations were established in the 19th and early 20th centuries and have not been updated to reflect northward production migration.

Supplementary Discussion: Honest Assessment of Limitations

Several features of the results warrant discussion that extends beyond the main text, particularly where null findings or methodological constraints may influence interpretation.

The Q5–Q1 trend differential. Although basis volatility trends upward in the highest migration-exposure quintile across four of five crops, the differential between Q5 and Q1 trend magnitudes is not statistically significant ($\tau = 0.10$, $p = 0.3900$). This result does not indicate that migration-exposed counties experience the same basis risk trajectory as unexposed counties; rather, it reflects the spatial coarseness of the 20-market AMS data, which span the entire continental United States but provide insufficient geographic resolution to detect cross-quintile divergence that may exist at finer scales. Elevator-level data from individual grain merchandisers would provide the granularity required to resolve this question, though such data are privately held and not currently available for academic research.

Instrumental variable identification. Constructing a valid instrument for the delivery gap presents fundamental challenges. Because the delivery gap is a deterministic function of county location and CME delivery-point coordinates—both of which are fixed—any instrument that predicts delivery-gap variation must predict county geography. County geography, in turn, correlates with land values, transportation infrastructure, proximity to processing facilities, and other economic fundamentals that independently affect basis behaviour. The exclusion restriction is therefore inherently difficult to satisfy in this setting. The primary soil-climate instrument failed first-stage relevance testing ($F \approx 0$), and an alternative terrain-based instrument, while achieving strong first-stage performance ($F = 1{,}801$), violated the exclusion restriction by directly predicting futures prices ($t = 61.13$, $p < 0.0001$). Given these constraints, OLS estimates with extensive controls constitute the primary specification, and readers should note the accompanying endogeneity caveat.

Binary versus continuous difference-in-differences. The binary DiD specification produces a null treatment effect ($p = 0.4298$), while the continuous specification using delivery-gap distance as treatment intensity yields a highly significant coefficient ($\beta = 0.568$, $p = 0.0001$). These results are not contradictory. The continuous specification is theoretically more appropriate in this context: there is no reason to expect a discontinuous jump in hedge-ratio effects at an arbitrary geographic boundary, and the delivery-gap mechanism operates through transportation costs that scale linearly with distance. The significant continuous result and null binary result together indicate a gradient rather than a threshold effect.

Welfare model decomposition. Ablation A10 demonstrates that freezing the delivery gap at 1990 values produces welfare estimates identical to the dynamic specification. This result indicates that the welfare framework as currently implemented does not isolate the incremental cost of ongoing migration from the level effect of pre-existing delivery-gap distances. The \$5.34 billion annual estimate reported in the main text should therefore be interpreted as the total cost of geographic mismatch between production and delivery infrastructure, of which the migration-attributable increment (\$0.11 billion per year) is estimated through a separate production-reallocation decomposition rather than through the welfare model directly.

Forecasting model performance. Seasonal naive and random walk benchmarks achieve lower RMSE than TFT-FULL on the held-out test period. This outcome is consistent with the properties of basis as a near-martingale process at weekly horizons, where strong seasonal patterns and mean-reverting dynamics favour simple benchmarks over complex architectures. The contribution of the TFT analysis is not point-forecast superiority but rather the demonstration—through SHAP attribution and controlled ablation—that geographic features carry non-redundant predictive information absent from all standard basis models.

These limitations, taken together, do not invalidate the delivery-gap mechanism documented in the main text. They do, however, underscore the need for expanded data infrastructure—including finer-grained AMS price reporting, public access to elevator-level bid data, and spatially explicit cash-price indices—that would enable the research community to measure basis risk with precision commensurate with its economic importance.

Supplementary References

Lim et al. 2021 Lim, B., Arık, S. Ö., Loeff, N. & Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* **37**, 1748–1764 (2021).