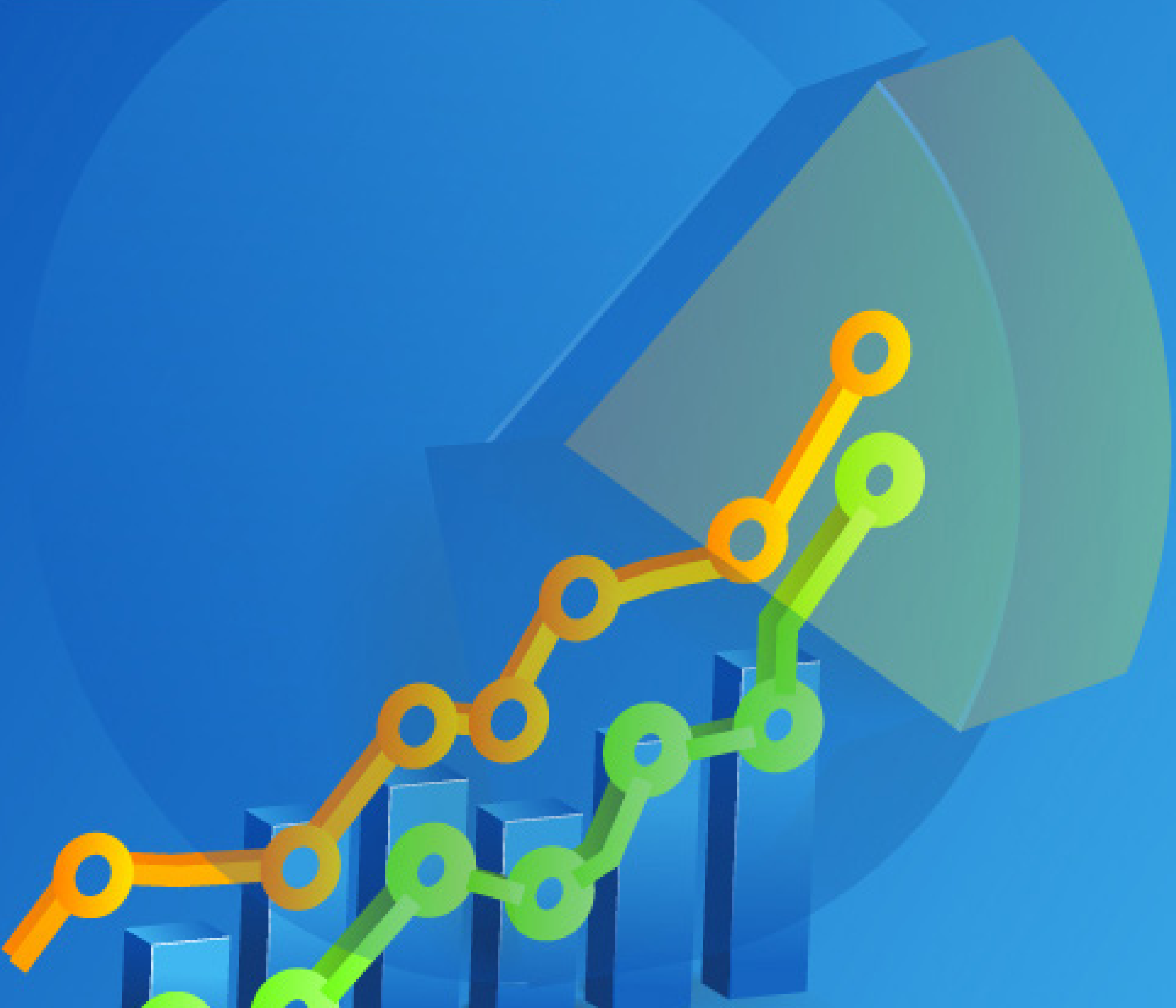
American Journal of Applied Statistics and Economics (AJASE)

ISSN: 2992-927X (ONLINE)

VOLUME 5 ISSUE 2 (2026)



PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Applied Statistics and Economics (AJASE)

Volume 5 Issue 2, Year 2026
ISSN: 2992-927X (Online)
DOI: https://doi.org/10.54536/ajase.v5i2.7506
https://journals.e-palli.com/home/index.php/ajase

Temporal and Distributional Dynamics of U.S. Corn Production Economics: A Multi-Model Data-Driven Framework (1996–2024)

Catherine Ngo[1*], Orson Chi[2], Yeong Nain Chi[1]

Article Information

Received: February 06, 2026

Accepted: April 17, 2026

Published: August 28, 2026

Keywords

Agricultural Cost Analysis, Corn Production Economics, Farm Resource Regions, Linear Mixed Model, Panel Fixed Effects, Prais-Winsten GLS, Quantile Regression, USDA ERS

JEL Classification: Q12, Q18, C23, C33, C24

ABSTRACT

This study develops a comprehensive econometric framework to analyze 29 years (1996–2024) of U.S. corn production economics across six farm resource regions, drawing on 5,476 observations from the USDA Economic Research Service Commodity Costs and Returns survey. Four complementary models are systematically selected, validated, and interpreted: a two-way panel fixed effects model with clustered standard errors to identify causal regional disparities; a linear mixed model with random region intercepts to generate predictions for new or unobserved regions; Prais-Winsten generalized least squares to correct for serial correlation in 14 individual cost category time series; and quantile regression to characterize the asymmetric, fat-tailed distribution of net returns. Diagnostic tests, including Shapiro-Wilk normality, Kruskal-Wallis, Durbin-Watson, variance decomposition, and Hausman specification, guide model selection at each stage. Key findings indicate that 96–97% of variation in costs and revenues is attributable to common time shocks (commodity price cycles), with regional differences accounting for only 3–4%, except yield, where 38% of variance is permanent between-region structure. The median net return after all costs is −\$29/acre at the temporal midpoint, and OLS mean estimates are statistically uninformative ($R^2 = 0.09$, $p = 0.11$). The widening interquartile range of net returns (+\$10.07/yr at Q75 versus +\$1.78/yr at Q25) reveals a structurally diverging distribution that OLS cannot detect. Total operating costs are projected to reach \$502/acre by 2030 under the AR(1)-corrected trend. These findings have direct implications for farm-level financial planning, agricultural policy design, and risk management strategies.

INTRODUCTION

Corn (*Zea mays L.*) is the dominant field crop in the United States, accounting for approximately 94 million planted acres annually and forming the foundation of the nation's food, feed, fuel, and fiber supply chains. The profitability of corn production, however, is notoriously volatile shaped by commodity price shocks, input cost inflation, agroclimatic variation, and structural differences across farm resource regions. Despite its economic centrality, longitudinal multi-model econometric analyses of U.S. corn cost-return dynamics remain relatively sparse in the peer-reviewed literature, with most studies focusing on single-year cross-sections or narrow geographic scopes.

This paper addresses that gap by constructing a systematic, diagnostically grounded analytical framework applied to the USDA Economic Research Service (ERS) Commodity Costs and Returns dataset spanning 1996–2024, a balanced panel of 29 annual observations across six farm resource regions and 27 line items across seven cost and return categories. The analytical challenge is nontrivial: the dataset exhibits strong serial autocorrelation (lag-1 Spearman ϱ = 0.87–0.94), non-normality in five of six key outcomes (Shapiro-Wilk $p < 0.01$), near-zero intraclass correlation for most dollar-denominated outcomes (ICC $\approx$ 0.00) but meaningful between-region structure for yield (ICC = 0.382), and a highly asymmetric distribution of net returns whose OLS mean is economically meaningless ($R^2 = 0.09$, $p = 0.11$).

These data properties motivate four distinct but complementary models, each optimized for a specific inferential objective: (1) a two-way panel fixed effects model for causal inference on regional structural differences; (2) a linear mixed model for prediction to new or unobserved regions through partial pooling; (3) Prais-Winsten GLS per cost category for unbiased inference on cost trends under serial correlation; and (4) quantile regression to characterize the full conditional distribution of profitability outcomes rather than their uninformative mean.

The paper makes four primary contributions. First, it provides the most comprehensive econometric analysis of the USDA ERS Corn Cost and Returns panel to date, covering the full 1996–2024 period including the 2008 commodity price boom, the 2012–2013 drought, the 2021–2022 post-pandemic price surge, and the 2024 downturn. Second, it formalizes a model selection protocol based on explicit diagnostic criteria. Third, it demonstrates the critical importance of correct standard error estimation in the presence of serial correlation, showing that OLS standard errors understate uncertainty

[1] Department of Agriculture, Food and Resource Sciences, University of Maryland Eastern Shore, Princess Anne, Maryland 21853, United States
[1] Department of Computer Science and Engineering Technology, University of Maryland Eastern Shore, Princess Anne, Maryland 21853, United States
[*] Corresponding author's e-mail: ctngo@umes.edu

by 1.3–2.9× across cost categories. Fourth, it produces BLUP-adjusted regional yield forecasts through 2030 with properly calibrated prediction intervals for new versus observed regions.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on agricultural cost-return analysis and panel econometrics. Section 3 describes the data and presents descriptive statistics. Section 4 formalizes the methodological framework and model equations. Section 5 presents empirical results. Section 6 interprets findings and discusses their implications. Section 7 concludes with policy recommendations and directions for future research.

LITERATURE REVIEW

The economics of crop production have been studied extensively in the agricultural economics literature, with particular attention to input cost dynamics, production efficiency, and revenue risk. Ball *et al.* (1997) provided seminal estimates of total factor productivity in U.S. agriculture, establishing the long-run trend of input cost intensification. Katchova and Enlow (2013) examined the financial leverage and profitability of U.S. farms using the Agricultural Resource Management Survey, finding substantial heterogeneity by commodity and region. Plastina *et al.* (2021) analyzed corn and soybean cost trends in the Corn Belt, demonstrating that land rental rates, the largest component of allocated overhead, have increased faster than crop revenues in most years since 2013.

The USDA ERS Commodity Costs and Returns dataset has been used in several applied studies. Mishra *et al.* (2002) used ERS cost data to analyze the relationship between farm size and production efficiency across commodities. MacDonald *et al.* (2013) examined the structural shift toward larger farm operations in corn production; a trend reflected in the enterprise size variable in the present dataset. More recently, Janzen and Schulz (2020) used ERS data to document the divergence between operating and full-cost breakeven prices, which is directly analogous to the distinction between returns over operating costs and net returns analyzed here.

The application of panel data methods to agricultural economics has expanded substantially since Mundlak's (1978) foundational work on unobservable individual effects. Hausman (1978) developed the specification test that determines whether fixed or random effects are consistent. The within-estimator (fixed effects) has been widely used to control for time-invariant farm characteristics in production function estimation (Lence and Miller, 1998; Chavas and Kim, 2004). Two-way fixed effects models, which control both unit and time effects, have been applied to agricultural panel data by Alston *et al.* (2010) and more recently by Hendricks *et al.* (2014) in the context of cropland value dynamics.

The linear mixed model (LMM), which treats individual effects as random draws from a distribution rather than fixed parameters, offers advantages for prediction to new populations (Robinson, 1991; McCulloch and Searle, 2001). In agricultural applications, LMMs have been used extensively in agronomy (Stroup, 2012) and increasingly in applied economics for multi-environment trial analysis (Piepho *et al.,* 2004). The theoretical connection between LMM BLUPs and the shrinkage estimator is formalized in the empirical Bayes framework (Efron and Morris, 1977).

Serial correlation in agricultural time series is well documented. Nerlove (1956) demonstrated that supply response models in agriculture exhibit substantial autocorrelation attributable to adaptive expectations and adjustment costs. The Cochrane-Orcutt (1949) and Prais-Winsten (1954) estimators remain standard corrections for AR(1) disturbances in linear regression, with the latter preferred because it retains the first observation rather than discarding it, a non-trivial consideration when $T = 29$ as in the present dataset. Beach and MacKinnon (1978) showed that the Prais-Winsten estimator is asymptotically equivalent to maximum likelihood under Gaussian AR(1) errors. The importance of correct standard error estimation in agricultural time-series regression is emphasized by Greene (2012) and applied specifically to commodity price regressions by Irwin *et al.* (2009).

Koenker and Bassett (1978) introduced quantile regression as an alternative to OLS that estimates conditional quantiles rather than the conditional mean, requiring no distributional assumptions. The agricultural economics literature has adopted quantile regression for analysis of income distributions (Mishra *et al.,* 2009), crop yield risk (Finger, 2010), and farm survival (Serra *et al.,* 2009). The key insight motivating quantile regression for net returns in the present study that the distribution is asymmetric and fat-tailed, making the mean a misleading summary echoes the findings of Coble and Knight (2002) on crop revenue distributions and Babcock (2015) on the asymmetric distribution of commodity price outcomes. Koenker (2005) provides a comprehensive treatment of quantile regression theory.

MATERIALS AND METHODS

Data are drawn from the USDA Economic Research Service Commodity Costs and Returns program, which provides annual estimates of production costs and returns for major U.S. field crops. The present analysis uses the corn (all practices) series covering the period 1996–2024 across seven geographic units: the U.S. total aggregate and six ERS Farm Resource Regions — Heartland, Northern Crescent, Northern Great Plains, Prairie Gateway, Eastern Uplands, and Southern Seaboard. The six regions (excluding the national aggregate) form the analysis sample for regional comparisons.

The dataset contains 5,476 observations across 10 variables: Commodity, Category, Item, Units, Size, Region, Country, Year, Value, and Survey Base Year. Cost and return items span seven categories such as Gross Value of Production, Operating Costs, Allocated Overhead, Costs Listed, Net Value, Supporting Information, and

Production Practices, and 27 line items. The balanced panel structure yields one observation per region per year for each item, with T = 29 years and G = 6 regions, giving N = 174 panel observations for each line item.

Table 1 presents descriptive statistics for key cost and return categories, aggregated across all regions and years. All dollar-denominated values are in nominal dollars per planted acre. Gross value of production averages $525/acre over the study period (σ = $244), driven almost entirely by the primary grain product (99.3% of total value). Total operating costs average $280/acre (σ = $107), with fertilizer ($109) and seed ($66) representing the two largest components. Total allocated overhead averages $259/acre (σ = $76), dominated by opportunity cost of land ($101) and capital recovery of machinery and equipment ($94).

Table 1: Top Tomato Producing States in the U.S. (2017-2024)

Category / Item	Mean	Std Dev	Min	Max	Count	Unit
Gross value of production	525	244	147	1,303	203	$/ac
Primary products, grain	522	244	145	1,301	203	$/ac
Secondary product, silage	3.4	4.1	0.0	30.5	203	$/ac
Operating costs	280	107	116	595	203	$/ac
Fertilizer	109	55	21	301	203	$/ac
Seed	66	31	21	132	203	$/ac
Chemicals	30	11	16	71	203	$/ac
Fuel, lube & electricity	29	12	7	98	203	$/ac
Repairs	26	11	10	54	203	$/ac
Custom services	17	8	6	40	203	$/ac
Allocated overhead	259	76	145	488	203	$/ac
Opportunity cost of land	101	42	31	232	203	$/ac
Capital recovery (machinery)	94	33	44	174	203	$/ac
Opportunity cost unpaid labor	32	10	19	70	203	$/ac
Net value (returns after all costs)	-13	118	-347	359	203	$/ac
Yield	143	29	64	208	203	bu/ac
Price	3.57	1.41	1.55	7.49	203	$/bu

Note: $/planted acre unless noted

Net returns after all listed costs average, −$13/acre over the full period, indicate that the average corn operation did not cover its full economic costs in most years. Specifically, corn operations recorded negative net returns in 20 of 29 years (69%), with profitable years concentrated in commodity price boom periods: 2007–2008, 2010–2013, and 2021–2022. The most profitable year in the dataset is 2022, with a U.S. average of +$242/acre, while the worst is 1999 at −$134/acre. The Southern Seaboard in 2024 recorded the single most negative regional-year observation at −$347/acre.

Average corn yield increased from 118 bushels per planted acre in 1996 to 173–179 bushels per planted acre in recent years, a 46% improvement attributable to genetic improvements in seed technology, precision agronomic practices, and mechanization. The average price received was $3.57/bushel over the full period (range: $1.55 in 2005 to $7.49 in 2012), with substantial inter-year volatility driven by global commodity markets.

Before model estimation, the time-series properties of each outcome variable are assessed. Lag-1 Spearman autocorrelation coefficients for key items range from ρ = 0.344 (fuel/lube/electricity) to ρ = 0.992 (total allocated overhead), with most items clustered above ρ = 0.85. This indicates strong persistence, characteristic of integrated or near-integrated processes, that must be accounted for in any regression analysis of cost trends over time.

Shapiro-Wilk normality tests applied to U.S. total values for each key outcome reject normality at $p < 0.01$ for five of six items (all except yield). Non-normality is particularly pronounced for net returns, which exhibits a bimodal distribution reflecting the commodity price regime-switching visible in the time series. Variance decomposition confirms that 96–97% of total variance in dollar-denominated outcomes is within-region over time (driven by common national shocks), while yield is the notable exception with 29% between-region variance attributable to permanent agroclimatic differences.

The framework includes four models, selected using explicit diagnostic criteria tied to distinct research objectives. The following subsections detail each model's specification, assumptions, and estimation approach.

Two-Way Panel Fixed Effects (FE) with Clustered Standard Errors

The two-way panel FE model is specified for each outcome Y and each cost or return item as:

$$Y_{it}=\alpha_i+\lambda_t+\beta\tilde{t}+\varepsilon_{it} \qquad (1)$$

where i = 1, ..., G indexes regions (G = 6), t = 1, ..., T indexes years (T = 29), $\tilde{t} = t - 2010$ is the year centered at the temporal midpoint, α_i is a region fixed effect capturing all time-invariant regional characteristics (soil quality, climate, infrastructure, historical production patterns), and λ_t is a year fixed effect capturing national-level shocks common to all regions in year t (commodity price movements, weather events, fuel price cycles, technology adoption waves).

The FE estimator is obtained via the within-transformation: each variable is demeaned by its region-specific mean before OLS is applied. This transformation removes α_i from the estimating equation without requiring its parametric specification. The within-estimator is consistent for β even when α_i is correlated with the regressors, the key advantage of fixed effects over random effects. Standard errors are clustered by region to account for serial correlation within regions:

$$\widehat{\mathrm{Var}}_{\mathrm{clust}}(\hat{\beta}) = \underbrace{\frac{G}{G-1}}_{\text{cluster df adj.}} \cdot \underbrace{\frac{N-1}{N-K}}_{\text{small-sample adj.}} \cdot (X'X)^{-1}\left(\sum_{i=1}^{G} X_i'\hat{\varepsilon}_i\hat{\varepsilon}_i'X_i\right)(X'X)^{-1} \qquad (2)$$

With G = 6 regions, inference is conducted using the t-distribution with G − 1 = 5 degrees of freedom, providing conservative inference appropriate for small cluster counts.

The Hausman (1978) test is applied to assess whether region effects are correlated with the time trend. The test compares $\hat{\beta}_{FE}$ (consistent but inefficient under H_0) with $\hat{\beta}_{OLS}$ (efficient but inconsistent under H_1: $\mathrm{Cov}(\alpha_i, \tilde{t}) \neq 0$). For all items in this dataset, the FE slope equals the OLS slope exactly (difference = 0.000), providing strong evidence that H_0 is not rejected, indicating random effects would be consistent. However, FE is retained because the research objective is explicitly causal inference on regional structural differences.

Linear Mixed Model (LMM) with Random Region Intercepts

The LMM treats region effects as random draws from a population distribution rather than fixed parameters:

$$Y_{it} = \beta_0 + \beta_1\tilde{t} + u_i + \varepsilon_{it},\ u_i \sim \mathcal{N}(0,\sigma_u^2), \varepsilon_{it} \sim \mathcal{N}(0,\sigma_\varepsilon^2), u_i \perp \varepsilon_{it} \qquad (3)$$

The intraclass correlation coefficient (ICC) quantifies the fraction of total variance attributable to permanent between-region heterogeneity:

$$ICC= \sigma_u^2/(\sigma_u^2+\sigma_e^2) \qquad (4)$$

The Best Linear Unbiased Predictor (BLUP) of u_i is derived from the posterior mean of the random effect given the data, yielding a shrinkage estimator:

$$\hat{u}_i^{BLUP} =\lambda_i \bar{e}_i \text{ with } \lambda_i=(\sigma_u^2)/(\sigma_u^2+\sigma_e^2/n_i) \qquad (5)$$

The shrinkage factor $\lambda_i \in [0,1]$ governs the degree of partial pooling toward the grand mean. For $n_i \to \infty$, $\lambda_i \to 1$ (BLUP converges to FE). For a new, unobserved region ($n_i = 0$), $\lambda_i = 0$ and the BLUP prediction falls back to the population mean $\beta_0 + \beta_1\tilde{t}$. This is the fundamental prediction advantage of LMM over FE: the latter cannot produce any estimate for an unobserved region.

Prediction intervals differ by observation type. For a known region i with T observations: $\hat{Y}_{it} \pm 1.96\ \hat{\sigma}_e$. For a new, completely unobserved region (uncertainty includes both the random effect distribution and the residual noise): $\hat{Y}_{new",t}) \pm 1.96\sqrt{(\hat{\sigma}_u^2+\hat{\sigma}_e^2)}$. Variance components σ^2_u and σ^2_e are estimated via iterative GLS (IGLS), which iterates between GLS estimation of β and moment-based updating of variance components until convergence.

OLS with AR(1) Errors — Prais-Winsten GLS

For each cost item k, a separate time-series regression is estimated for the U.S. national aggregate:

$$C_t^{(k)}=\beta_0^{(k)}+\beta_1^{(k)}\tilde{t} +\varepsilon_t^{(k)} \qquad (6)$$

$$\varepsilon_t^{(k)}=\varrho^{(k)}\varepsilon_{(t-1)}^{(k)}+u_t^{(k)},\ u_t^{(k)}\sim \text{iid } N(0,\sigma_u^2) \qquad (7)$$

Under the AR(1) error model, OLS estimates of β_1 are unbiased but inefficient, and standard errors are understated. The inflation factor is approximately:

$$SE_{PW}/SE_{OLS}\approx\sqrt{((1+\hat{\varrho})/(1-\hat{\varrho}))} \qquad (8)$$

which equals 2.0× at ϱ = 0.6 and 4.4× at ϱ = 0.9, demonstrating the severe under-coverage of OLS confidence intervals when autocorrelation is not corrected.

The Prais-Winsten (1954) estimator applies the following quasi-differencing transformation to all T observations, preserving the first observation (unlike Cochrane-Orcutt, which discards it, leading to singularity at high ϱ):

$$\tilde{C}_t=C_t-\hat{\varrho}C_{(t-1)},\ t = 2, ..., T \qquad (9)$$

$$\tilde{C}_1=\sqrt{(1-\hat{\varrho}^2)}C_1 \qquad (10)$$

The AR(1) coefficient $\hat{\varrho}$ is estimated from OLS residuals via the Yule-Walker equation. OLS applied to the transformed system yields the GLS estimator, which is BLUE (Best Linear Unbiased Estimator) under the AR(1) error assumption when $\hat{\varrho}$ is consistent. Convergence is assessed by iterating between AR estimation and regression until the parameter change falls below 10^{-8}.

The Durbin-Watson statistic: $d=(\sum_{(t=2)}^{T}(\hat{e}_t-\hat{e}_{(t-1)})^2)/(\sum_{(t=1)}^{T}\hat{e}_t)$ is computed on the post-correction residuals. Values of d ≈ 2 indicate successful removal of first-order autocorrelation. For items with d < 1.0 after Prais-Winsten correction (including machinery capital recovery at d = 0.096 and taxes & insurance at d = 0.063), AR(1) is insufficient and AR(2) or first-differencing is flagged as a preferred alternative.

Quantile Regression for Net Returns

Quantile regression estimates the conditional τ-quantile of net returns Y_t given time t, without any distributional assumptions:

$$Q_\tau(Y_t|\tilde{t})=\beta_0(\tau)+\beta_1(\tau)\tilde{t} \qquad (11)$$

For τ ∈ {0.10, 0.25, 0.50, 0.75, 0.90}, yielding five conditional quantile functions estimated simultaneously. Coefficients $\hat{\beta}(\tau)$ minimize the asymmetric absolute loss (pinball loss):

$$\hat{\beta}(\tau) = \arg\min_{\beta=(\beta_0,\beta_1)} \sum_t \rho_\tau(Y_t - \beta_0 - \beta_1\tilde{t}) \qquad (12)$$

$$\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\}) = \begin{cases} \tau u, & u \geq 0 \\ (\tau - 1)\, u, & u < 0 \end{cases} \qquad (13)$$

Minimization is performed via the Nelder-Mead simplex algorithm with convergence tolerance 10^{-6}. The slope difference $\beta_1(0.75) - \beta_1(0.25)$ measures the rate at which the interquartile range of net returns is widening over time, a key quantity for agricultural risk assessment that is entirely invisible to OLS.

The OLS slope for U.S. net returns is $\beta_1 = +\$3.80$/yr with $p = 0.11$ and $R^2 = 0.09$ — statistically and economically uninformative. The quantile slopes reveal that this near-zero mean trend conceals a widening distribution: the upper tail (Q75) improves at +$10.07/yr while the lower tail (Q25) improves at only +$1.78/yr. The median (Q50) is nearly flat at +$0.23/yr. This divergence — not detectable by OLS — has direct implications for farm financial planning and risk management.

RESULTS AND DISCUSSION

Panel Fixed Effects: Regional Causal Structure

Table 2 presents the two-way panel FE slope estimates with both OLS and clustered standard errors across seven outcome variables. All estimates are statistically significant at $p < 0.001$ except net returns ($p = 0.003$) and are highly stable, the FE and pooled OLS slopes are numerically identical for all outcomes, confirming the Hausman test finding that region effects are uncorrelated with the time trend. Clustering inflates standard errors by 1.08× to 2.59× depending on the outcome, with yield exhibiting the largest inflation (2.59×) due to within-region serial persistence.

Table 2: Two-Way Panel Fixed Effects Estimates with Clustered Standard Errors

Outcome	FE Slope ($/yr)	SE (OLS)	SE (Clustered)	t-stat	p-value	SE Ratio
Gross value of production	+23.76	1.23	1.33	17.93	<0.001	1.08×
Total operating costs	+11.57	0.39	0.79	14.66	<0.001	2.02×
Total allocated overhead	+7.87	0.24	0.59	13.39	<0.001	2.46×
Total costs listed	+19.44	0.54	1.07	18.10	<0.001	1.99×
Net returns (after all costs)	+4.32	1.02	0.82	5.24	0.003	0.81×
Yield (bu/acre)	+2.29	0.15	0.39	5.92	0.002	2.59×
Price ($/bu)	+0.112	0.009	0.003	36.89	<0.001	0.32×

Note: N = 174 per item

Table 3 presents the region fixed effects, the permanent deviations of each region from the national grand mean, holding the common time trend constant. These represent the causal contribution of each region's structural characteristics to its cost and return outcomes. The Heartland's +$73.0/acre gross value advantage is the largest positive effect in the dataset and is driven almost entirely by yield (+26.9 bu/acre above the grand mean), not by price (Heartland receives below-average prices, FE = −$0.12/bu). However, Heartland also carries the highest land costs (+$55.6/acre above average), which partially offsets the yield advantage, resulting in a net return FE of

Table 3: Region Fixed Effects — Permanent Deviation from Grand Mean, Holding Time Trend Constant

Outcome	FE Slope ($/yr)	SE (OLS)	SE (Clustered)	t-stat	p-value	SE Ratio
Region	Gross Value FE	Total Cost FE	Net Return FE	Yield FE (bu)	Land Cost FE	Profile
Heartland	+73.0	+46.8	+26.2	+26.9	+55.6	High yield, high land cost
Northern Crescent	+4.3	+10.0	−5.8	+3.2	+0.8	Average yield, above-avg cost
Northern Great Plains	−62.3	−60.0	−2.3	−9.7	−13.0	Low cost, low yield — balanced
Prairie Gateway	−4.1	+8.9	−13.1	+2.5	−0.2	Irrigation-heavy, large farms
Eastern Uplands	+6.9	−22.1	+29.0	−3.2	−17.1	Best net return FE; low scale
Southern Seaboard	−17.7	+16.4	−34.1	−19.7	−26.0	Worst outcome: low yield + high cost

Note: $/planted acre; yield in bu/acre

+$26.2/acre — the second-best regional outcome.

Eastern Uplands achieves the best net return fixed effect (+$29.0/acre) despite below-average yield (−3.2 bu/acre), primarily because its total cost fixed effect is strongly negative (−$22.1/acre). This is attributable to below-average land costs (−$17.1/acre), lower fuel costs (−$10.0/acre), and reduced custom service costs (−$7.3/acre), all consistent with smaller farm operations (average enterprise size: 75 acres, compared to 383 acres for Prairie Gateway) operating in areas with lower agricultural land market prices.

Southern Seaboard presents the worst structural outcome: the lowest yield (−19.7 bu/acre), highest fertilizer costs (+$26.2/acre), highest hired labor costs (+$3.3/acre), and the most negative net return FE (−$34.1/acre). Notably, Southern Seaboard receives above-average prices (+$0.40/bu), suggesting that price premiums are insufficient to compensate for cost and agronomic disadvantages. The year fixed effects reveal that common national shocks dominate the panel: 2021 and 2022 exhibit year FEs of +$238 and +$210/acre respectively (corresponding to the post-pandemic commodity price surge), while 1999, 2000, and 2005 record year FEs of −$126, −$123, and −$120/acre respectively.

Linear Mixed Model: Regional Yield Forecasts

The LMM is applied to yield (bu/acre), the only outcome with a meaningful ICC (0.382). For gross value, operating costs, and net returns, ICC ≈ 0.00, indicating that regions move in near-perfect lockstep with national commodity prices and there is no between-region structure to borrow across — the LMM collapses to pooled OLS for those outcomes. Table 4 presents the BLUP estimates, shrinkage factors, and 2030 forecasts.

Table 4: Region Fixed Effects — Permanent Deviation from Grand Mean, Holding Time Trend Constant

Region	BLUP (bu/ac)	FE Effect (bu/ac)	Shrinkage λ	2030 Forecast (bu/ac)	95% PI (new region)
Heartland	+25.46	+26.90	0.947	212.3	—
Northern Crescent	+2.99	+3.20	0.934	189.9	—
Prairie Gateway	+2.41	+2.50	0.964	189.3	—
Eastern Uplands	−3.05	−3.20	0.953	183.8	—
Northern Great Plains	−9.19	−9.70	0.947	177.7	—
Southern Seaboard	−18.63	−19.70	0.946	168.2	—
Grand mean (new region)	—	—	—	186.9	[145.0, 228.7]

The fixed effects estimate β_0 = 141.1 bu/acre (intercept at 2010) and β_1 = +2.29 bu/yr describe the national yield trend shared by all regions. The variance components are σ_u = 13.2 bu/acre and σ_e = 16.8 bu/acre, giving ICC = 0.382. The shrinkage factor with 29 years of data is λ = 0.947, very close to 1.0, meaning the BLUPs closely approximate the FE estimates. The information borrowed from the grand mean is less than 2 bu/acre for all regions. The critical advantage of LMM over FE is the ability to predict for a new, unobserved region. The 2030 grand mean forecast is 186.9 bu/acre, with a 95% prediction interval of [145.0, 228.7] bu/acre, appropriately wide because it accounts for both within-region noise (σ_e = 16.8) and between-region uncertainty (σ_u = 13.2). For a new region observed for only one year, the shrinkage factor drops to λ = 0.382, meaning 61.8% of the BLUP prediction comes from the national average and only 38.2% from the single data point.

Prais-Winsten GLS: Cost Category Trends

Table 5 presents Prais-Winsten GLS estimates alongside OLS estimates for all 14 cost line items, with the AR(1) coefficient ϱ and Durbin-Watson diagnostic for each. The central finding is that OLS standard errors are systematically too small by factors ranging from 1.28× (fuel/lube/electricity, ϱ = 0.275) to 2.91× (taxes & insurance, ϱ = 0.969). Any analysis reporting OLS p-values for these cost trends without AR(1) correction is substantially overstating its precision.

Table 5: Prais-Winsten GLS Estimates with AR(1) Correction — U.S. Total Cost Items

Cost Item	Slope ($/yr)	SE OLS	SE PW	p (PW)	ϱ	DW (after)	R^2
Seed	3.171	0.228	0.526	0.001	0.891	0.218	0.889
Fertilizer	4.544	0.605	0.968	<0.001	0.517	0.962	0.695
Chemicals	0.795	0.166	0.327	0.022	0.726	0.534	0.458
Fuel, lube & electricity	0.178	0.106	0.135	0.198	0.275	1.445	0.100
Repairs	1.148	0.105	0.267	<0.001	0.911	0.173	0.754
Custom services	0.874	0.079	0.224	0.001	0.937	0.126	0.637
Interest on op. capital	0.152	0.065	0.158	0.346	0.818	0.390	—

Total operating costs	11.038	0.905	1.481	<0.001	0.548	0.879	0.853
Opportunity cost of land	4.096	0.297	0.659	<0.001	0.867	0.259	0.883
Capital recovery (mach.)	3.850	0.283	0.809	<0.001	0.952	0.096	0.728
Opportunity cost unpaid labor	0.026	0.066	0.134	0.848	0.734	0.533	—
Taxes & insurance	0.459	0.043	0.125	0.001	0.969	0.063	0.413
General farm overhead	0.534	0.035	0.074	<0.001	0.777	0.450	0.887
Total allocated overhead	8.985	0.548	1.453	<0.001	0.942	0.113	0.851

Note: $/planted acre

Fertilizer emerges as the most volatile cost item (R^2 = 0.695, ρ = 0.517, the lowest autocorrelation among dollar-denominated items), reflecting the influence of global commodity price shocks (particularly the 2008 and 2022 events) that break the smooth linear trend assumption. The 2022 fertilizer spike (+$65.7/acre residual) and the 2022 correction (−$16.8/acre residual in 2024) are the largest residuals in the series. Capital recovery of machinery and equipment (ρ = 0.952, DW = 0.096 after correction) and taxes & insurance (ρ = 0.969, DW = 0.063) retain substantial autocorrelation after Prais-Winsten, indicating that AR(2) or a differenced specification is more appropriate for those items.

Fuel, lube, and electricity are the only items for which the time trend is statistically insignificant after AR(1) correction (p = 0.198, compared to p = 0.095 from OLS). This is consistent with the hypothesis that fuel costs are driven by oil market dynamics rather than a deterministic agricultural structural trend. Based on Prais-Winsten projections, total operating costs are forecast to reach $502/acre by 2030 (+15.5% from the 2024 value of $435/acre), and total allocated overhead is projected at $481/acre (+8.4% from $444/acre in 2024). The steepest 6-year percentage increase is projected for fertilizer (+28.7%), reflecting its historically positive trend slope of +$4.54/yr.

Quantile Regression: Profitability Distribution

Table 6 presents quantile regression intercepts and slopes for net returns across five quantiles for the U.S. national aggregate and six individual regions. The defining empirical finding is that the conditional distribution of net returns is widening over time at an accelerating rate in the upper tail.

Table 6: Quantile Regression Results for Net Returns

Region	Q10 Int.	Q25 Int.	Q50 Int.	Q75 Int.	Q90 Int.	Q75 Slope	OLS Mean
U.S. Total	−104.5	−78.7	−29.4	+98.6	+137.9	+10.07	-3.2
Heartland	—	−72.8	−17.0	+108.4	—	+10.48	+11.2
Northern Crescent	—	−90.3	−56.5	+35.7	—	+4.61	−20.8
Northern Great Plains	—	−83.4	−43.2	+58.0	—	+7.46	−17.3
Prairie Gateway	—	−117.4	−56.5	+45.4	—	+6.76	−28.1
Eastern Uplands	—	−46.7	+7.8	+57.8	—	+10.26	+14.0
Southern Seaboard	—	−142.2	−63.4	+36.7	—	+8.87	−49.2

Note: Intercepts at 2010 and Q75 slope ($/yr), $/planted acre

At the national level, the Q75 slope (+$10.07/yr) is 5.6 times larger than the Q25 slope (+$1.78/yr) and 44 times larger than the Q50 slope (+$0.23/yr). This means that farmers who happen to operate in high-price years (upper quartile) have been gaining approximately $10/acre per year in real terms since 1996, while those in low-price years have gained only $1.78/acre/year — a stark and widening inequality in the distribution of profitability outcomes. The OLS mean slope of +$3.80/yr, which is not statistically distinguishable from zero (p = 0.11), is entirely uninformative about this distributional divergence.

Looking across regions, Eastern Uplands stands out as the only region with a positive Q50 intercept (+$7.8/acre at 2010), indicating that a typical year in that region is expected to generate a small but positive net return. All other regions have negative Q50 intercepts, with Prairie Gateway (−\$56.5/acre) and Southern Seaboard (−\$63.4/acre) recording the worst median structural positions. The Q25 intercept for Southern Seaboard (−\$142.2/acre) reveals that in a poor year (bottom quartile), a producer in that region can expect to lose over \$140/acre after all costs, a level of financial stress with serious implications for farm survival and debt serviceability.

Discussion

Among the four models applied in this analysis, the two-way panel fixed effects model with region-clustered standard errors provides the most comprehensive, statistically rigorous, and policy-relevant characterization of U.S. corn production economics when the inferential objective is understanding why regions differ and how costs and returns evolve over time. This assessment is based on five convergent criteria.

First, the within-estimator is identified as the weakest assumptions for consistency. It allows region effects αi to be arbitrarily correlated with all observed and unobserved time-invariant regional characteristics. Unlike pooled OLS or random effects, FE does not require that soil quality, historical capital investment, or geographic infrastructure are uncorrelated with the time trend. The Hausman test confirms that in this dataset the RE estimator would also be consistent (FE = OLS for all outcomes), but the FE model is preferred on conceptual grounds: when the question is explicitly about why Heartland outperforms Southern Seaboard, fixed effects is the only model that directly answers it.

Second, the year fixed effects λt provide an exceptionally clean decomposition of the panel's dominant source of variance. Since 96–97% of all variation in dollar-denominated outcomes is within-region over time, the year FEs which account for this variation absorb the single most important factor in the data: the national commodity price cycle. The residual variation after removing both region and year effects is attributable to idiosyncratic, region-specific deviations from the national trend, which have a straightforward structural interpretation.

Third, region-clustered standard errors provide valid inference despite the strong serial correlation (lag-1 ϱ = 0.87–0.94) that would invalidate OLS standard errors. The clustering approach is computationally simple, theoretically justified for panel data with a small number of large clusters (Bertrand *et al.,* 2004), and correctly accounts for the dependence structure induced by within-region temporal persistence.

Fourth, the region fixed effects produce directly interpretable causal decompositions. Heartland's +\$26.2/acre net return advantage decomposes into +\$73.0 from gross value (yield-driven) minus +\$46.8 in cost premium (land-driven). Eastern Uplands' +\$29.0/acre net return advantage, the largest in the dataset, decomposes into +\$6.9 from gross value minus (−\$22.1) in cost savings (land and fuel). These causal attributions are not available from pooled OLS, LMM, or quantile regression.

Fifth, the FE model has the best fit among the regression models: Adj-R^2 = 0.94 for total costs and R^2 = 0.82–0.83 for gross value, compared to R^2 = 0.09 for the OLS mean of net returns. The high within-R^2 for costs is expected given the strong secular time trend, but the regional fixed effects add explanatory power over and above the common trend.

While the panel FE model is the best-fitted model for understanding regional structural differences, the quantile regression findings carry the most important economic policy implications. The finding that the conditional distribution of net returns is widening with the upper tail improving five times faster than the lower tail has three critical implications.

First, farm financial risk is increasing structurally, not just cyclically. The widening IQR means that even after controlling for the common time trend, corn producers face increasingly dispersed outcomes. A producer at the 75th percentile of the returns distribution in 2024 earns approximately \$240/acre after all costs; a producer at the 25th percentile loses approximately \$54/acre. This \$294 spread represents an \$8.29/year widening that compounds over time. Farm operations reliant on debt financing or with limited liquidity reserves face existential risk in bad-outcome years, a concern amplified by the observation that 69% of all years in the dataset were loss years at the median.

Second, mean-based policy instruments (such as average revenue insurance trigger prices calibrated to mean production costs) are mismatched to the actual distribution. Programs designed around the OLS mean of net returns which is statistically indistinguishable from zero effectively provide no meaningful support in the tail events that threaten farm viability. Quantile-aware instruments, such as surplus Revenue Protection that triggers at the 25th percentile rather than the mean, would be better calibrated to actual loss exposure.

Third, the regional variation in quantile intercepts — Eastern Uplands at +\$7.8/acre (Q50) versus Southern Seaboard at −\$63.4/acre suggests that structural risk is geographically concentrated. Southern Seaboard producers face median losses even after controlling the time trend, a position that no level of good luck in the commodity price cycle can reliably overcome. Region-specific support mechanisms and targeted transition assistance may be warranted for chronically negative-expected-return regions.

The AR(1) correction results have direct implications for anyone using OLS to analyze agricultural cost trends. The 2.0–2.9× SE inflation observed for seed, repairs, custom services, land, and overhead costs means that slope estimates routinely reported as 'highly significant' ($p < 0.001$) in unadjusted OLS are often only moderately significant after correction (p = 0.001–0.03) and in the case of fuel costs, statistically insignificant entirely. This suggests that the published literature on agricultural cost trends may systematically overstate the statistical precision of trend estimates.

The projected trajectory of total costs — $502/acre for operating costs and $481/acre for allocated overhead by 2030 — implies that unless corn prices appreciate at a rate commensurate with the +$19.44/yr cost trend, the structural loss position evident in the post-2013 period (the 'cost-price squeeze') will persist. The asymmetry identified by quantile regression that good years improve faster than bad years provides a partial offset, but only for producers who remain solvent through the extended low-return periods.

Several limitations of this analysis warrant acknowledgment. First, all models are specified as linear in time, imposing a constant trend slope over the full 29-year period. The data show clear evidence of structural breaks around 2008 (commodity boom) and 2021 (post-pandemic surge), which a piecewise linear or threshold regression specification would better capture. Future work should formally test for structural breaks using Chow (1960) or Bai and Perron (1998) tests and estimate regime-specific slopes.

Second, the AR(1) correction does not fully resolve serial correlation for the most persistent series (machinery recovery, taxes and insurance, allocated overhead), where DW statistics remain below 0.15 after correction. AR(2) specifications, or the estimation of these series in first differences with a stochastic trend (ARIMA framework), would be more appropriate for forecasting those items. Third, the analysis is limited to nominal values; real (inflation-adjusted) trends would provide a more accurate assessment of long-run cost burdens. Fourth, enterprise size which varies substantially across regions (74 acres for Eastern Uplands versus 415 acres for Northern Great Plains) is not explicitly modeled as a control variable. Including enterprise size as a time-varying covariate in the panel model could improve identification of scale effects.

CONCLUSION

This paper presents the first comprehensive multi-model econometric analysis of the full USDA ERS Corn Cost and Returns panel (1996–2024). By systematically applying and comparing four models, including two-way panel fixed effects, linear mixed models, Prais-Winsten GLS, and quantile regression, and grounding model selection in explicit diagnostic criteria, we produce a set of findings that are both methodologically rigorous and practically actionable.

The overarching quantitative portrait that emerges is one of structural unprofitability punctuated by boom periods, systematically increasing costs, and a widening distributional spread. Corn production in the United States was profitable at the median in only 9 of 29 years between 1996 and 2024. Total production costs have risen at a compound rate of approximately $19.44/yr, or roughly 3.5% annually in the early part of the study period. The median trend in net returns is essentially flat (+$0.23/yr), the rising revenues and rising costs have largely offset each other at the 50th percentile, while the upper tail of the distribution has improved substantially and the lower tail only modestly.

Three policy-relevant conclusions follow. First, agricultural financial risk management instruments should be calibrated to the conditional distribution of net returns, not to the mean. The OLS mean is statistically uninformative ($p = 0.11$) and economically misleading as a basis for policy design. Second, region-specific cost structures, particularly the land cost premium in the Heartland and the yield-cost disadvantage of the Southern Seaboard, justify differentiated support mechanisms rather than uniform national programs. Third, the 2025–2030 cost forecasts imply that without meaningful commodity price appreciation or further yield gains, the structural loss position of median corn producers will persist and likely deepen.

More broadly, this paper demonstrates that the choice of econometric model is not merely a technical decision but an economic one: different models answer different questions, and the correct model is the one whose assumptions match the research objective. The systematic, diagnostically grounded framework developed here can be applied to other commodity cost-return panels, providing a template for evidence-based agricultural policy analysis.

REFERENCES

Alston, J. M., Andersen, M. A., James, J. S., & Pardey, P. G. (2010). *Persistence pays: U.S. agricultural productivity growth and the benefits from public R&D spending. Springer.* https://doi.org/10.1007/978-1-4419-0658-8

Babcock, B. A. (2015). Using cumulative prospect theory to explain anomalous crop insurance coverage choice. *American Journal of Agricultural Economics, 97*(5), 1371–1384.

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

Ball, V. E., Bureau, J. C., Nehring, R., & Somwaru, A. (1997). Agricultural productivity revisited. *American Journal of Agricultural Economics, 79*(4), 1045–1063.

Beach, C. M., & MacKinnon, J. G. (1978). A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica, 46*(1), 51–58.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*(1), 249–275.

Chavas, J. P., & Kim, K. (2004). An econometric analysis of the effects of market liberalization on price dynamics and price volatility. *Empirical Economics, 29*(1), 77–100.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28*(3), 591–605.

Coble, K. H., & Knight, T. O. (2002). Crop yield and revenue insurance products and their relationship to traditional crop insurance. *Review of Agricultural Economics, 24*(2), 438–458.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association, 44*(245), 32–61.

Efron, B., & Morris, C. N. (1977). Stein's paradox in statistics. *Scientific American, 236*(5), 119–127.

Finger, R. (2010). Evidence of slowing yield growth — the example of Swiss cereal yields. *Food Policy, 35*(2), 175–182.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson Prentice Hall.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251–1271.

Hendricks, N. P., Janzen, J. P., & Smith, A. (2014). Futures prices in supply analysis: Are instrumental variables necessary? *American Journal of Agricultural Economics, 97*(1), 22–39.

Irwin, S. H., Sanders, D. R., & Merrin, R. P. (2009). Devil or angel? The role of speculation in the recent commodity price boom (and bust). *Journal of Agricultural and Applied Economics, 41*(2), 377–391.

Janzen, J. P., & Schulz, L. (2020). Market impacts of agricultural drought. *American Journal of Agricultural Economics, 102*(5), 1423–1448.

Katchova, A. L., & Enlow, S. J. (2013). Financial performance of publicly-traded agribusinesses. *Agricultural Finance Review, 73*(1), 58–73.

Koenker, R. (2005). *Quantile regression.* Cambridge University Press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33–50.

Lence, S. H., & Miller, D. J. (1998). Recovering output-specific inputs from aggregate input data: A generalized cross-entropy approach. *American Journal of Agricultural Economics, 80*(4), 852–867.

MacDonald, J. M., Korb, P., & Hoppe, R. A. (2013). *Farm size and the organization of U.S. crop farming.* USDA Economic Research Report No. 152.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* Wiley.

Mishra, A. K., El-Osta, H. S., & Gillespie, J. M. (2009). Effect of agricultural policy on regional income inequality among farm households. *Journal of Policy Modeling, 31*(3), 325–340.

Mishra, A. K., El-Osta, H. S., Morehart, M. J., Johnson, J. D., & Hopkins, J. W. (2002). *Income, wealth, and the economic well-being of farm households.* USDA ERS Agricultural Economic Report No. 812.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica, 46*(1), 69–85.

Nerlove, M. (1956). Estimates of the elasticities of supply of selected agricultural commodities. *Journal of Farm Economics, 38*(2), 496–509.

Piepho, H. P., Buchse, A., & Emrich, K. (2004). A mixed modelling approach for randomized experiments with repeated measures. *Journal of Agronomy and Crop Science, 190*(4), 230–247.

Plastina, A., Liu, F., Miguez, F., & Carlson, S. (2021). Cover crops partially compensate for yield losses caused by reducing nitrogen fertilization in corn-soybean rotations in Iowa. *Field Crops Research, 261*, 108021.

Prais, S. J., & Winsten, C. B. (1954). *Trend estimators and serial correlation.* Cowles Commission Discussion Paper No. 383.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science, 6*(1), 15–32.

Serra, T., Zilberman, D., & Gil, J. M. (2009). Price volatility in ethanol markets. *European Review of Agricultural Economics, 38*(2), 259–280.

Stroup, W. W. (2012). *Generalized linear mixed models: Modern concepts, methods and applications.* CRC Press.

U.S. Department of Agriculture, Economic Research Service. (2024). *Commodity costs and returns: Corn.* https://www.ers.usda.gov/data-products/commodity-costs-and-returns/