The Effect of China's GMO Ban on Crop Yields

Yuansen Li[1], Tor N. Tolhurst*[2], and Matthew Gammans[3]

[1]PhD Candidate, Purdue University

[2]Assistant Professor, Michigan State University

[3]Assistant Professor, North Dakota State University

Abstract

Genetically engineered (GE) maize varieties, introduced in the 1990s, have been widely adopted in the Americas and have delivered large productivity gains. In contrast, China has prohibited their commercial cultivation. Understanding the consequences of this ban requires more than an estimate of average productivity losses. For both producers and consumers, what matters for income, food security, and welfare is not just the average, but the entire distribution of yields — especially the risk of very low outcomes. However, recovering a credible counterfactual distribution is essentially impossible with standard methods, which are designed for means or a few quantiles and cannot capture the evolution of the full distribution over time. We address this challenge by applying a distributional synthetic control approach to province-level data from 1970–2019, using the United States, Argentina, and Brazil as donor regions. This method recovers the entire dynamic trajectory of the counterfactual distribution, allowing us to trace how the technology would have shifted not just the mean but also dispersion, tail risks, and stability over time without imposing strong functional form assumptions. For China's three largest maize-producing provinces—together accounting for roughly half of national production—preliminary estimates indicate median shortfalls of –18.9%, –29.8%, and –15.2%, respectively, with even larger penalties at the 80th percentile (–20.8%, –28.9%, and –17.0%). We also find increases of 17.5%, 16.1%, and 11.1% in the probability that yields fall below 90% of their mean. These effects are statistically significant under conservative inference procedures. The findings suggest that prohibiting GE maize has imposed productivity and stability losses on a scale comparable to major technological transformations in modern agricultural history.

*Corresponding author: tolhurs2@msu.edu. The authors received no financial support for the research, authorship, and/or publication of this article. The authors declare that they have no relevant financial or non-financial interests to disclose.

1 Introduction

Agricultural productivity differs enormously across countries. For maize—the world's most widely cultivated cereal—yields in the most productive regions can be several times higher than in the least productive (Gollin et al., 2014). These gaps matter for farm incomes, food security, and trade balances, and a substantial share reflects differences in access to and adoption of modern crop-breeding technologies (Evenson and Gollin, 2003; Alston and Pardey, 2014). The Green Revolution transformed production in much of Asia and Latin America (Gollin et al., 2021); more recently, genetic engineering, introduced in the mid-1990s, has delivered substantial productivity gains in the Americas (Qaim and Zilberman, 2003; Areal et al., 2013).

Regulatory approaches to genetically engineered (GE) crops—often referred to as genetically modified organisms (GMOs) in regulatory and policy discourse—have diverged sharply since their commercial introduction in the mid-1990s. The United States, Argentina, and Brazil approved GE maize early and moved rapidly toward near-universal adoption. In contrast, many other countries, in both the Global North and South, have prohibited or heavily restricted their cultivation. Bans have been maintained in Peru, Venezuela, and Ecuador; long-standing prohibitions historically covered food crops in India, Kenya, and Tanzania; and several African countries, including Zambia, have upheld planting bans on precautionary grounds even in the face of food shortages. These divergent policies have created a rare setting in which regions with broadly comparable growing conditions operate under fundamentally different technological regimes—a divergence especially salient for maize, given its global importance, large observed yield gaps, and the wide transferability of GE-driven productivity gains across similar agro-climatic zones.[1]

China's regulatory trajectory exemplifies the restrictive approach to GE crops. Beginning in the mid-1980s, the government invested heavily in biotechnology research, developing domestic breeding programs and expanding laboratory and field-trial capacity. In 1997, China approved Bt cotton for commercial cultivation, which was rapidly adopted and became one of the earliest large-scale GE crop successes outside the Americas. Yet no GE grain crop received approval for domestic cultivation over the next two decades. Imports of GE products were tightly regulated, with approval

[1]For example, Northeast China (Heilongjiang, Jilin) has a temperate continental climate with summer rainfall patterns closely resembling the U.S. Corn Belt (Iowa, Illinois), including comparable growing-season lengths, precipitation during planting and flowering stages, and frost windows. The North China Plain (Henan, Shandong) shares climatic features with Argentina's Pampas, particularly in temperature profiles and rainfall distribution.

processes that were lengthy and uncertain—for example, GE alfalfa, approved for planting in the United States in 2011, was only cleared for import into China in 2023. By 2019, GE cotton remained the only widely planted GE crop in the country, even as maize demand expanded and yield growth became essential to offset limited arable land. Although official statements periodically endorsed GE technology as a strategic tool for food security, the policy ban persisted until late 2023, when the Ministry of Agriculture and Rural Affairs approved the first GE maize and soybean varieties for commercial planting.[2]

A large literature documents the productivity and efficiency gains from GE crops in adopting countries. Using experimental plot data, Chavas et al. (2014) show that GE maize adoption shifts the entire yield distribution, with impacts extending beyond the mean. In a commercial production setting, Lusk et al. (2018) likewise find effects on average yields, variability, and tail outcomes. Scheitrum et al. (2020) broaden the analysis to multiple crops and traits in a multi-country panel, documenting substantial aggregate production gains. Most recently, Hansen and Wingender (2023) exploit crop-by-country variation in the timing of trait availability in a triple-difference design to identify heterogeneous yield impacts. Taken together, this work establishes that GE crops have sizeable effects on both mean yields and higher moments of the yield distribution. In agriculture, these distributional shifts are economically salient: an upward shift in the upper tail raises the likelihood of surplus and export potential, while compression of the lower tail reduces the risk of shortfalls and lowers crop insurance costs. What remains elusive, however, is a counterfactual estimate of how the entire yield distribution would have evolved under a full-scale GE ban.

Recovering such a distribution is essentially impossible with standard methods, which are designed to estimate means or a few quantiles and cannot replicate the full shape of the distribution. In adoption-based designs, the challenge is compounded by the fact that variation in adoption intensity is rarely exogenous: regions with higher adoption shares often differ systematically in pest pressure, input access, or management, creating potential bias from endogenous adoption. In shift–share settings, this generates a fundamental trade-off: large pre-period share differences yield more identifying variation but heighten susceptibility to endogeneity, whereas small differences mitigate that risk but identify only marginal changes among adopters—such as moving from 90% to

[2]For example: `https://www.economist.com/china/2023/02/23/is-chinas-attitude-to-genetically-modified-crops-changing`.

91% adoption—rather than the effect of moving from a full-scale ban to near-universal adoption. Because control units are typically drawn from partial adopters rather than true non-adopters, these designs cannot directly recover the distribution that would have prevailed under a full-scale ban.

We address this challenge by applying the distributional synthetic control (DSC) method of Gunsilius (2023) to province-level maize yields in China from 1970–2019, using GE-adopting regions in the United States, Argentina, and Brazil as the donor pool. DSC extends the conventional synthetic control framework (Abadie et al., 2010, 2015) by matching the full pre-treatment yield distribution—rather than just its mean—using the Wasserstein-2 distance. This approach produces counterfactual distributions from which we can analyze central tendency, dispersion, tail risks, and other distributional features. Crucially, it also allows us to recover the *entire dynamic trajectory* of the counterfactual yield distribution, making it possible to see not just whether, but exactly how, the technology would have reshaped yields over time in the absence of the ban.

Our setting requires adapting the DSC estimator to cases where the unit-level, time-conditional distribution is not directly observed. In agriculture, we typically have only one yield observation per unit-year, rather than a rich panel of micro-level outcomes. Fortunately, there is a well-established literature in crop insurance and agricultural risk analysis on how to estimate these distributions from limited annual data. We follow this approach to construct the necessary pre-treatment yield distributions, and we incorporate a clustered jackknife procedure to ensure that the uncertainty from these estimation steps is propagated through to all downstream inference.

We also make a methodological contribution by proposing a pooled, time-invariant donor-weight DSC estimator that uses all pre-treatment years jointly. Sparse and time-invariant weights are central to the logic of synthetic control, ensuring that the counterfactual is constructed from a consistent combination of donor units rather than a shifting set each year. While one could average the weights from separate year-by-year problems, the resulting solution is generally not the same as the solution to the pooled problem, and it can fail to match the joint pre-treatment distribution as well as possible. Our approach directly solves the pooled optimization, producing weights that are both stable and efficient, reducing overfitting and improving interpretability. This addresses an open problem left by Gunsilius (2023) and can be applied wherever the DSC estimator is appropriate. More broadly, the policy evaluation literature recognizes that interventions can shift not just the

mean but higher moments and shape parameters of the outcome distribution (Bitler et al., 2006; Firpo et al., 2009); by improving how proper donor weights are estimated for full distributions, our approach could help expand this literature's ability to assess the complete impacts of policies and technologies.

Our estimates indicate that the GE maize ban substantially reduced yields in China's major producing provinces. For the largest three—Heilongjiang, Jilin, and Inner Mongolia, together accounting for roughly half of national output—median yields in 2019 would have been 18.9%, 29.8%, and 15.2% higher, respectively, in the counterfactual without the ban, with even larger proportional gains at the 80th percentile (20.8%, 28.9%, and 17.0%). For the next provinces, median and 75th-percentile yields would have been higher in Henan (7.4% and 27.9%) and Hubei (38.2% and 42.3%). Shandong's estimates also imply yield penalties, but are not statistically different from zero. Beyond shifts in central tendency, the ban also increased downside risk in most provinces. In 2019, the probability of yields falling below 90% of their mean was higher by 17.5%, 16.1%, and 11.1% in Heilongjiang, Jilin, and Inner Mongolia, respectively, and by 12.6% in Henan and 10.5% in Hubei. The estimate for Shandong is slightly negative and not statistically significant. These changes indicate that, alongside reducing typical yields, the ban also made low-yield outcomes more likely in five of the six major producing provinces. Because the observed and counterfactual distributions continue to diverge over time, maintaining the ban would be expected to widen these productivity and stability gaps even further.

Our counterfactual estimates are at the high end of, and in some cases exceed, yield gains reported in the GE maize literature. Prior studies—ranging from experimental trials to multi-country panel designs—typically find increases of 8–30%. The relatively large impacts here reflect our comparison of sustained zero adoption with the global technology frontier, capturing the cumulative benefits of decades of trait improvements and hybrid adaptation rather than marginal short-run changes.

This paper makes three contributions. First, it provides the first quantitative assessment of the distributional effects—across the full dynamic trajectory—of China's long-standing ban on GE maize, documenting how the policy has altered yields and stability in one of the world's largest maize producers. Second, it extends the policy evaluation literature by showing how the DSC framework can be adapted to recover counterfactual yield distributions in settings where only one

outcome is observed per unit-year, with appropriate inference that accounts for distribution estimation. Third, it offers a methodological contribution of independent interest by developing a pooled, time-invariant donor-weight estimator that stabilizes synthetic control weights over long pre-treatment periods, reducing overfitting and improving interpretability. Together, these contributions demonstrate the importance of examining the entire distribution of agricultural outcomes—and its evolution over time—when evaluating technological change, particularly in environments where risk and stability matter as much as mean productivity.

The remainder of the paper is organized as follows. Section 2 outlines the DSC methodology and describes the adaptations required for our application, including the pooled weight estimator. Section 3 presents the data sources and construction of the provincial yield distributions. Section 4 reports the empirical results, including province-level counterfactual distributions, measures of dispersion, and tail probabilities. Section 5 concludes with implications for China's biotechnology policy and for the evaluation of agricultural technologies in settings where risk and stability are central policy concerns.

2 Empirical Model

2.1 Distributional Synthetic Control Estimator

The synthetic control method (SCM) estimates treatment effects by constructing a weighted average of untreated units that approximate the treated unit's outcomes in the absence of treatment (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). In the classical setting, the goal is to recover the unobserved counterfactual outcome for treated unit j post-treatment:

$$\delta_{jt} = Y_{jt} - Y_{jt}^{N} \tag{1}$$

where Y_{jt} is the observed outcome and Y_{jt}^{N} is the unobserved counterfactual; that is, the outcome that would have occurred in the absence of treatment. Following the conventional notation, the treated unit is indexed by $j = 1$ and the J donor units by $j = 2, \ldots, J + 1$. The pre-treatment period is denoted $\mathcal{T}_0 = \{1, \ldots, T_0\}$.

The synthetic control estimator constructs a counterfactual outcome for the treated unit using

a convex combination of untreated units:

$$\widehat{Y}_{1t}^{N} = \sum_{j=2}^{J+1} w_j \cdot Y_{jt}. \tag{2}$$

The weights $\mathbf{w} = (w_j)_{j=2}^{J+1}$ are chosen to minimize pre-treatment differences between the treated unit and its synthetic counterpart, typically by matching outcomes or covariates over the pre-treatment period $\mathcal{T}_0$. Identification relies on the assumption that the same weights would have continued to approximate the treated unit's outcomes in the absence of treatment; that is, the relationship between the treated unit and the donor pool would have remained stable over time had the treatment not occurred.

This framework is naturally extended to the distributional setting by replacing the scalar outcome Y_{jt} with its distribution, denoted $F_{Y_{jt}}$. The target of inference becomes the counterfactual distribution for the treated unit, $F_{Y_{1t}^N}$. Just as classical SCM constructs a synthetic control to match pre-treatment outcomes, distributional synthetic control (DSC) aims to construct a synthetic distribution that matches the treated unit's distribution over the pre-treatment period. And just as the classical SCM problem centers on finding the weights that generate a counterfactual for mean outcomes, the core challenge in the distributional case is to find weights that credibly approximate the counterfactual distribution.

Formally, the distributional treatment effect is defined as the difference between the treated and counterfactual cumulative distribution functions (CDFs):

$$F_{\delta_{1t}} = F_{Y_{1t}} - F_{Y_{1t}^N}, \tag{3}$$

where $F_{Y_{jt}}$ denotes the CDF of outcomes for unit j in year t, and $f_{Y_{jt}}$ denotes the corresponding probability density function (PDF), when it exists. The counterfactual CDF $F_{Y_{1t}^N}$ is constructed using a convex combination of donor units:

$$\widehat{F}_{Y_{1t}^N} = \sum_{j=2}^{J+1} w_j \cdot \widehat{F}_{Y_{jt}}. \tag{4}$$

Unlike in the classical setting, we never observe distributions directly—only their realizations—so

the right-hand side necessarily involves estimates, as indicated by the hat.

However, as in the classical case, the goal remains the same: to choose weights w_j that closely match the treated unit's distribution over the pre-treatment period. The key contribution of Gunsilius (2023) is to show that this problem can be effectively solved by minimizing the Wasserstein-2 (W_2) distance between the treated unit and a convex combination of distributions from the donor pool. In doing so, this approach overcomes several limitations of alternative methods for comparing distributions.

First, as Gunsilius shows, parametric strategies (such as a mixture of Normals) can obscure important distributional features when averaged. For example, taking the convex combination of two bimodal mixtures can create a unimodal distribution, even though none of the original units are unimodal. This smoothing effect can erase meaningful heterogeneity and lead to misleading counterfactuals. While this issue can be mitigated by using nonparametric estimators or restricting attention to distributions with bounded support, a second limitation remains more fundamental.

Pointwise metrics—such as the integrated squared differences between densities or Kullback–Leibler divergence—perform poorly when the support of the treated and donor distributions differs, regardless of how those distributions are estimated. These measures either become infinite or undefined. For example, the integrated squared difference between $\mathcal{U}(-1,0)$ and $\mathcal{U}(0+\varepsilon,1)$ for a small $\varepsilon > 0$ is infinite, despite the two distributions being arbitrarily close in location. Further, pointwise metrics assign the same divergence between $\mathcal{U}(0,1)$ and $\mathcal{U}(1+\varepsilon,2)$ as they do between $\mathcal{U}(0,1)$ and $\mathcal{U}(2,3)$, despite the latter being farther away. These measures fail to account for relative distance and instead treat any pair of non-overlapping supports as equally dissimilar.

In contrast, the Wasserstein-2 distance remains well-defined and sensitive to the geometry of distributions, even when their supports do not overlap. The squared W_2 distance between two distributions F and G, with quantile functions $F^{-1}(\tau)$ and $G^{-1}(\tau)$, is given by:

$$W_2^2(F,G) = \int_0^1 \left|F^{-1}(\tau) - G^{-1}(\tau)\right|^2 d\tau, \tag{5}$$

where $\tau \in [0,1]$ indexes the quantile level. The W_2 criterion defines distance between distributions in terms of differences across quantiles.

Intuitively, the Wasserstein-2 distance measures how much "work" is required to reallocate

probability mass from one distribution to another, accounting for both the amount of mass shifted and the distance it must be moved. Rooted in optimal transport theory, minimizing this distance corresponds to solving for the most efficient reallocation of mass needed to transform one distribution into the other. Because it compares entire quantile functions, the Wasserstein-2 distance preserves features such as skewness, heavy tails, and multimodality, making it particularly well suited for applications where shape—not just location or scale—is policy-relevant.

To estimate the counterfactual distribution, Gunsilius (2023) solves for weights that minimize the squared Wasserstein-2 distance between the treated unit and a convex combination of donor distributions in each pre-treatment year. That is, let

$$\mathbf{w}_t \in \Delta^{J-1} \equiv \left\{ \mathbf{w}_t \in \mathbb{R}^J : w_{jt} \geq 0 \, \forall j, \sum_{j=2}^{J+1} w_{jt} = 1 \right\}$$

denote the vector of time-varying weights, where Δ^{J-1} is the $(J-1)$-dimensional unit simplex. For a given time $t \leq T_0$, Gunsilius (2023) solves:

$$\min_{\mathbf{w}_t \in \Delta^{J-1}} \int_0^1 \left(\widehat{F}^{-1}_{Y_{1t}}(\tau) - \sum_{j=2}^{J+1} w_{jt} \cdot \widehat{F}^{-1}_{Y_{jt}}(\tau) \right)^2 d\tau. \tag{6}$$

This problem is convex in $\mathbf{w}_t$ and admits a unique global minimum. It solves the synthetic control problem one year at time.

However, this one-year-at-a-time strategy raises its own issue. Estimating a separate set of weights for each pre-treatment year, indicated by w_{jt}, requires an additional step to aggregate them into a single time-invariant vector. Gunsilius (2023) addresses this by taking a simple average across [3]

$$\hat{w}^*_j = \frac{1}{T_0} \sum_{t<T_0} \hat{w}^*_{jt}. \tag{7}$$

The problem is that averaging across individually estimated solutions does not, in general, recover the solution to the pooled problem; because the Wasserstein-2 objective is nonlinear, the average of the year-specific minimizers is not itself the minimizer of the average loss. Furthermore, averaging can produce weight vectors that are dense or dispersed, which undermines the sparsity essential to

[3] It is equivalent to solving a multi-objective optimization problem. Our method in equation (11) solves the multi-objective problem by uniformly weighting all objectives.

guard against overfitting in synthetic control (Chen, 2023). Intuitively, this is because averaging noisy weights across a finite number of years allows unstable donors to enter the solution, diluting signal with year-specific noise.

We address this problem by stacking the time-indexed vectors into a single pooled optimization problem. The solution is a single set of time-invariant weights that minimizes the average discrepancy across all pre-treatment years. Let the stacked vector of estimated quantile functions for the treated unit be:

$$\widehat{\mathbf{F}}_1^{-1}(\tau) = \begin{pmatrix} \widehat{F}_{Y_{11}}^{-1}(\tau) \\ \widehat{F}_{Y_{12}}^{-1}(\tau) \\ \vdots \\ \widehat{F}_{Y_{1T_0}}^{-1}(\tau) \end{pmatrix}, \tag{8}$$

where $\widehat{\mathbf{F}}_1^{-1}(\tau) \in \mathbb{R}^{T_0}$ collects the estimated quantile functions $\widehat{F}_{Y_{1t}}^{-1}(\tau)$ across all pre-treatment periods. Then denote the $T_0 \times J$ matrix stacking the corresponding donor quantile functions as:

$$\widehat{\mathbf{F}}_{\mathcal{D}}^{-1}(\tau) = \begin{pmatrix} \widehat{F}_{Y_{21}}^{-1}(\tau) & \widehat{F}_{Y_{31}}^{-1}(\tau) & \dots & \widehat{F}_{Y_{J+1,1}}^{-1}(\tau) \\ \widehat{F}_{Y_{22}}^{-1}(\tau) & \widehat{F}_{Y_{32}}^{-1}(\tau) & \dots & \widehat{F}_{Y_{J+1,2}}^{-1}(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{F}_{Y_{2T_0}}^{-1}(\tau) & \widehat{F}_{Y_{3T_0}}^{-1}(\tau) & \dots & \widehat{F}_{Y_{J+1,T_0}}^{-1}(\tau) \end{pmatrix}. \tag{9}$$

where $\mathcal{D} = \{2, ..., J+1\}$ indexes the set of donor units. This stacked structure allows us to frame the problem as a single optimization over all pre-treatment periods, which enforces time-invariant weights by construction. Then we can estimate the counterfactual by solving the pooled optimization problem:

$$\min_{\mathbf{w} \in \Delta^{J-1}} \left\| \widehat{\mathbf{F}}_1^{-1}(\tau) - \widehat{\mathbf{F}}_{\mathcal{D}}^{-1}(\tau)\mathbf{w} \right\|_2^2, \tag{10}$$

where $\mathbf{w} = (w_j)_{j=2}^{J+1} \in \Delta^{J-1}$ is the vector of time-invariant donor weights and Δ^{J-1} is the unit simplex.

This formulation pools information across all pre-treatment years and solves a single least-squares problem for each quantile level $\tau \in [0, 1]$. Compared to averaging year-by-year solutions, this approach enforces time-invariant weights directly within the minimization, accommodates variation across time, and guards against overfitting by concentrating weights on donors that consis-

tently provide good matches.

We follow Gunsilius (2023) in implementing the optimization problem by discretizing the quantile index $\tau \in [0,1]$ with a set of $M = 500$ quantile levels $\{\tau_m\}_{m=1}^M$ drawn from a uniform distribution. Whereas Gunsilius (2023) takes these draws from $\mathcal{U}(0,1)$, we draw from $\mathcal{U}(\varepsilon, 1-\varepsilon)$ where $\varepsilon = 0.025$. This adjustment reflects the practical limitations of crop yield data, where only a single observation per unit-year pair is available and the pre-treatment period spans just 27 years (1970–1996). Under these conditions, the empirical quantile function may be poorly defined near the extremes.[4] The tolerance ε is chosen to correspond roughly to the largest quantile jump $(1/T_0)$, ensuring that all sampled quantiles lie within the observed support of the outcome distribution.

The discrete version of the pooled optimization problem is:

$$\min_{\mathbf{w} \in \Delta^{J-1}} \frac{1}{M} \sum_{m=1}^{M} \left\| \widehat{\mathbf{F}}_1^{-1}(\tau_m) - \widehat{\mathbf{F}}_{\mathcal{D}}^{-1}(\tau_m)\mathbf{w} \right\|_2^2. \tag{11}$$

This formulation stacks each of the M pre-treatment quantile-matching problems into a single least squares criterion. The result is a standard quadratic program in $\mathbf{w}$, subject to simplex constraints, which can be efficiently solved using a convex optimization solver such as `quadprog` or `cvxpy`.

Once the optimal weights $\widehat{\mathbf{w}}^*$ are estimated, we construct the time-conditional counterfactual distribution for post-treatment periods by applying the optimal weights to the donor units' quantile functions:

$$\widehat{F}^{-1}_{Y^N_{1t}}(\tau) = \sum_{j=2}^{J+1} \widehat{w}_j^* \cdot \widehat{F}^{-1}_{Y_{jt}}(\tau), \quad \text{for } t > T_0. \tag{12}$$

The treatment effect on the whole distribution is then computed as the discrete version of the Wasserstein-2 distance:

$$\widehat{W}_2\Big(\widehat{F}_{Y_{1t}}, \widehat{F}_{Y^N_{1t}}\Big) = \left\{ \frac{1}{M} \sum_{m=1}^{M} \big[\widehat{F}^{-1}_{Y_{1t}}(\tau_m) - \widehat{F}^{-1}_{Y^N_{1t}}(\tau_m)\big]^2 \right\}^{1/2}. \tag{13}$$

This estimates the treatment effect at the whole outcome distribution. Alternatively, we may be

[4] Robustness checks show the results are not sensitive to the exact choice of ε in $\mathcal{U}(\varepsilon, 1-\varepsilon)$.

interested in the heterogeneous treatment effect (treatment effect at quantiles):

$$\widehat{F}^{-1}_{Y_{1t}}(\tau_m) - \widehat{F}^{-1}_{Y^N_{1t}}(\tau_m), \tag{14}$$

which allows us to look at the impact at a given quantile τ. In empirical applications, we also visualize these effects by plotting the corresponding PDFs for the treated and synthetic control units to highlight treatment-induced changes in the shape of the distribution.

2.2 Estimating Empirical Quantile Functions

The DSC estimator requires plug-in estimates of the empirical quantile functions, $\widehat{F}^{-1}_{Y_{jt}}(\tau)$. Following standard practice in the crop insurance literature, we model each unit-year (j,t) as having its own conditional distribution of yields. Specifically, we treat the observed time series $\{y_{js}\}_{s=1}^{T_0}$ as a sample from the distribution of Y_{jt} and use it to estimate the time-conditional density $\widehat{f}_{Y_{jt}}$ and corresponding quantile function $\widehat{F}^{-1}_{Y_{jt}}(\tau)$. We proceed in three steps: (1) detrending, (2) heteroskedasticity adjustment, and (3) kernel density smoothing.

Step 1: Time-Conditional Mean

To incorporate systematic improvements in average crop productivity over time, we estimate a flexible time-series model for each unit using a two-knot linear spline, similar to the method used by the USDA Risk Management Agency (RMA) in its rating procedure (c.f., Harri et al., 2011).[5] For each unit j, yields are modeled as:

$$y_{jt} = \alpha_j + \beta_j t + \gamma_j (t - \kappa_{lower})_+ + \eta_j (t - T_0)_+ + e_{jt}, \tag{15}$$

where $(x)_+ \equiv max(0, x)$ and $\kappa_{lower} \in \{s+1, s+2, \ldots, T_0 - s\}$ denotes the interior knot selected via 10-fold cross-validation, with $s = 5$ defining the support edge to guard against overfitting. The upper knot is fixed at T_0 for all units. This flexible specification allows for changes in trend within the sample period. In particular, the interior knot allows for slower yield growth early in the sample. To further protect against overfitting, we impose a weak monotonicity condition requiring

[5] Our approach is similar to the RMA procedure for yield detrending but differs in two respects: (1) we do not apply Huber or biweight adjustments, and (2) we do not impose spatial priors when selecting knot locations.

the fitted trend to be non-decreasing, as it is well-established that yields have increased over time.

Step 2: Time-Conditional Variance

We also account for changes in the variability of the yield distribution, which has increased over time alongside rising mean yields, though not necessarily in direct proportion (Harri et al., 2011; Ker and Tolhurst, 2019). For each unit j, we estimate:

$$\log(\hat{e}_{jt}^2) = \phi_0 + \phi_1 \log(\hat{y}_{jt}) + \phi_2 \cdot \mathbb{1}\{t > T_0\} \cdot \log(\hat{y}_{jt}) + v_{jt} \tag{16}$$

where $\hat{e}_{jt}$ and $\hat{y}_{jt}$ are the residuals and fitted values from the detrending step. The post-treatment period interaction term allows the variance of yields to evolve differently after the introduction of GE seeds.

Because estimates of the heteroskedasticity coefficients (ϕ_1, ϕ_2) can be unstable in finite samples, we bound the correction to avoid implausibly steep increases in variance. Specifically, we follow Harri et al. (2011) in capping the implied exponent governing the volatility adjustment at 2—that is, we restrict $\hat{\phi}_1 \leq 4$ and, if $\hat{\phi}_1 + \hat{\phi}_2 > 4$, we replace $\hat{\phi}_2$ with the adjusted value $\hat{\phi}_2' = \hat{\phi}_2 - (\hat{\phi}_1 - 4)$.[6]

To estimate the time-conditional yield distribution for year s, we apply the heteroskedasticity adjustment to the residuals from all pre-treatment years $t = 1, \ldots, T_0$. That is, for each residual $\hat{e}_{jt}$, we construct:

$$\tilde{y}_{jst} = \hat{y}_{js} + \hat{e}_{jt} \left(\frac{\hat{y}_{js}}{\hat{y}_{jt}} \right)^{\frac{1}{2}\left(\hat{\phi}_1 + \hat{\phi}_2 \cdot \mathbb{1}\{t > T_0\}\right)}, \tag{17}$$

where $\tilde{y}_{jst}$ represents a draw from the time-conditional distribution of yields for unit j in year s. This transformation rescales the residuals to reflect the changing volatility implied by shifts in the the predicted mean yield. The collection $\{\tilde{y}_{jst}\}_{t=1}^{T_0}$ forms the estimated time-conditional distribution for year s.

[6]The exponent in the heteroskedasticity adjustment is $\frac{1}{2}(\hat{\phi}_1 + \hat{\phi}_2 \cdot \mathbb{1}\{t > T_0\})$. When this exponent equals 1, residuals are rescaled linearly in the predicted mean, implying a constant coefficient of variation. Values above (below) 1 imply increasing (decreasing) relative variability as mean yields rise.

Step 3: Time-Conditional Densities and Quantile Functions

To obtain the time-conditional quantile functions used in the DSC procedure, we apply kernel smoothing to the adjusted yield data. For each unit j and year s, we estimate the empirical quantile function $\widehat{F}^{-1}_{Y_{js}}(\tau)$ using the adjusted sample $\{\tilde{y}_{jst}\}_{t=1}^{T_0}$, which reflects the time-conditional yield distribution for unit j in year s.

We begin by estimating a kernel density over a fixed grid of yield values $\{X_k\}_{k=1}^{G}$, where $G = 1000$. The density is estimated using a Gaussian kernel with a unit- and time-specific bandwidth. To ensure comparability across units and years, we estimate the baseline bandwidth on standardized residuals. Let $\hat{\sigma}_{\tilde{y}_{jst}}$ denote the sample standard deviation of the draws $\{\tilde{y}_{jst}\}_{t=1}^{T_0}$ in year s, centered at the conditional mean $\hat{y}_{js}$. Then the standardized residuals are: $\tilde{e}_{jst} = (\tilde{y}_{jst} - \hat{y}_{js})/\hat{\sigma}_{\tilde{y}_{jst}}$ for $t = 1, \ldots, T_0$. We pool these standardized residuals $\tilde{e}_{jst}$ across units and use maximum likelihood cross-validation to select an optimal bandwidth on this scale, denoted h_0. The resulting bandwidth is then rescaled by the local standard deviation in each unit-year to obtain the final smoothing parameter, $h_{js} = h_0 \cdot \hat{\sigma}_{\tilde{y}_{jst}}$.

To mitigate boundary bias in near-zero yields, we apply the reflection method. Specifically, the KDE is applied to the reflected sample $\{-\tilde{y}_{jst}, \tilde{y}_{jst}\}$, and the estimated density is then symmetrized and truncated to the positive domain: $\widehat{f}_{js}(y) = 2 \cdot \widehat{f}^{\mathrm{KDE}}_{js}(y)$ for $y > 0$. The empirical CDF is constructed via cumulative summation over the grid:

$$\widehat{F}_{js}(X_k) = \sum_{\ell=1}^{k} \widehat{f}_{js}(X_\ell) \cdot \Delta x, \tag{18}$$

where Δx is the grid spacing. We rescale the CDF to ensure it reaches unity at the upper end of the grid, preserving valid distributional properties.

To retrieve the estimated quantile function, we use the same set of $M = 500$ quantile levels $\{\tau_m\}_{m=1}^{M}$ drawn earlier from the trimmed uniform distribution $\mathcal{U}(\varepsilon, 1-\varepsilon)$, with $\varepsilon = 0.025$, as described in the previous subsection. Letting $\widehat{F}_{js}(X_k)$ denote the estimated CDF value at grid point X_k, we construct the quantile function by linear interpolation:

$$\widehat{F}^{-1}_{Y_{js}}(\tau_m) = g\left(\widehat{F}_{js}(X_k), X_k\right)(\tau_m), \quad \text{for } m = 1, \ldots, M, \tag{19}$$

where $g(\cdot)$ denotes a linear interpolation function that maps quantile levels to yield values. In practice, this is implemented using the `approxfun()` routine in `R`. These estimated quantile functions $\widehat{F}^{-1}_{Y_{js}}(\tau_m)$ are then used as inputs to the DSC procedure described above.

2.3 Inference via Jackknife Resampling

We conduct inference using the clustered jackknife of Hansen (2022). This provides a straightforward, nonparametric way to assess the variability of the estimated weights and, by extension, the counterfactual distributions and treatment effects. We cluster by year to allow for spatial dependence across units. That is, for each $t \in \mathcal{T}_0$, we drop year t from the pre-treatment period and re-run the entire estimation procedure—including detrending, heteroskedasticity correction, kernel density estimation, and quantile function construction—using the leave-one-out sample $\mathcal{T}_0 \setminus \{t\}$.

This procedure generates a sequence of jackknife estimates of the donor weights:

$$\widehat{\mathbf{w}}^{(-t)} = \arg \min_{\mathbf{w}^{(-t)} \in \Delta^{J-1}} \left\{ \frac{1}{M} \sum_{m=1}^{M} \left\| \widehat{\mathbf{F}}_1^{-1,(-t)}(\tau_m) - \widehat{\mathbf{F}}_{\mathcal{D}}^{-1,(-t)}(\tau_m) \mathbf{w}^{(-t)} \right\|_2^2 \right\}, \tag{20}$$

where $\widehat{\mathbf{F}}_1^{-1,(-t)}(\tau_m)$ and $\widehat{\mathbf{F}}_{\mathcal{D}}^{-1,(-t)}(\tau_m)$ are the quantile function stacks excluding year t. These jackknife weights are then used to construct counterfactual quantile functions and associated statistics of interest. Let $\hat{\theta}$ denote a statistic of interest—distributional treatment effects, but also scalar functionals of quantile functions, such as tail mass, interquartile range, etc.—and $\hat{\theta}^{(-t)}$, its analog computed using the jackknife sample excluding year t. Then the jackknife variance of $\hat{\theta}$ is:

$$\widehat{\mathrm{Var}}\left[\hat{\theta}\right] = \frac{T_0 - 1}{T_0} \sum_{t=1}^{T_0} \left(\hat{\theta}^{(-t)} - \bar{\theta}^{(\cdot)} \right)^2, \tag{21}$$

where $\bar{\theta}^{(\cdot)} = \frac{1}{T_0} \sum_{t=1}^{T_0} \hat{\theta}^{(-t)}$ is the average across all jackknife estimates.

This approach captures the sampling variation in the weights induced by the finite pre-treatment period and propagates it through to all downstream quantities of interest, including counterfactual distributions, quantile treatment effects, and functionals thereof. While somewhat computationally intensive, the procedure is fully nonparametric and avoids strong assumptions about the estimator's sampling distribution.

Table 1: Summary of Data and Coverage by Country

Country	Source	Adoption	Years	Units
Panel A. Treatment Units (GE Ban)				
China	National Bureau of Statistics of China	NA	1970 - 2019	Province (6 of 23)
Panel B. Donor Pool (GE Adoption)				
Argentina	National Directorate of Agriculture	1997	1970 - 2019	Province (15 of 23)
Brazil	Brazilian Institute of Geography and Statistics	2006	1974 - 2019	State (25 of 26)
United States	US Department of Agriculture NASS QuickStats	1996	1970 - 2019	State (41 of 50)

Notes: This table summarizes the primary sources, adoption years, temporal coverage, and subnational units for agricultural data used in the analysis. The time series spans 1970–2019 for all countries to ensure comparability, based on the earliest year with available data from China. For the United States and Argentina, earlier data exist but are not used in the analysis to preserve a consistent time frame across countries. "Adoption Date" refers to the year of national policy implementation or the earliest confirmed commercial adoption at scale.

3 Data

We construct a panel dataset of subnational crop production for China, Argentina, Brazil, and the United States, covering the period 1970–2019. Table 1 summarizes the sources, coverage, and administrative units used in the analysis. China serves as the treatment case, where GE crops are not approved for commercial cultivation. The remaining countries form the donor pool, each having adopted GE crop varieties at different points in time. To ensure comparability, all yield measures are converted to a common unit of kilograms per hectare (kg/ha), standardizing across differences in reporting conventions across countries.

For the treatment group, we restrict the analysis to six major Chinese agricultural provinces: Heilongjiang, Jilin, Inner Mongolia, Shandong, Henan, and Hubei. These provinces are among the country's largest producers of maize and other staple crops, and together account for a substantial (60–70%) share of China's total grain output. Heilongjiang and Jilin are part of the Northeast China Plain, a key maize-growing region with large-scale mechanized farms. Inner Mongolia also contributes significantly to maize and soybean production under semi-arid conditions. These three provinces alone typically account for 40–50% of China's total maize output. Shandong and Henan, located in the North China Plain, are high-output regions with intensive double-cropping systems.

Hubei, while more diversified in crop mix, is a major inland agricultural province with significant maize and rice cultivation. To preserve these regional differences in agro-climatic conditions, production systems, and policy environments, we estimate the DSC model separately for each province. This avoids pooling across structurally distinct regions, which could wash out localized effects and obscure meaningful heterogeneity in the response to the treatment.

To ensure comparability across countries, the analysis period begins in 1970 (the earliest year with consistent data available from China) even though earlier historical data exist for Argentina and the United States. Subnational coverage varies by country due to differences in reporting and data quality; for each country, we include only the set of provinces or states with continuous and internally consistent time series. Adoption dates refer to the year of national approval or earliest confirmed commercial use of GE crops at scale. All data are compiled from official statistical agencies as detailed in Table 1.

Two features of the data require accommodation in the estimation procedure. First, production data for Brazil are unavailable prior to 1974, which limits the pre-treatment window relative to other countries. Rather than dynamically adjusting the donor pool or imputing missing values, we rely on the estimated time-conditional yield distribution, $\hat{f}_{js}(y)$, which underlies our implementation of the DSC. This allows us to simply project the distribution backward for Brazil despite the missing early years. And any additional uncertainty this adds to the estimation process will be captured by the jackknife re-sampling inference.

Second, GE crop adoption is staggered, with the United States adopting in 1996, Argentina in 1997, and Brazil in 2006. To address this, we exclude data from transition periods (i.e, 1997 for Argentina and 1997–2007 for Brazil) so they do not influence the estimated pre- or post-treatment distributions. Because we estimate treatment effects separately for each treated unit, the procedure avoids the complications associated with staggered treatment timing in panel estimators, such as those identified by de Chaisemartin and d'Haultfoeuille (2020).

4 Preliminary Results

This section presents preliminary results in three steps. We first estimate time-conditional maize yield distributions for each province from 1970 to 2019. Next, we compare the observed pre- and

post-treatment distributions under the GE ban to counterfactual distributions generated using the distributional synthetic control method, which approximate yields had the ban not been in place since 1996. Finally, we use these counterfactual distributions to construct test statistics for key features of the distribution—central tendency, dispersion, and shape—providing a quantitative basis for evaluating how the ban affected both average yields and production risk.[7]

4.1 Time-Conditional Yield Distributions

Figure 1 presents the estimated time-conditional yield density and temporal process for Inner Mongolia, the third-largest maize-producing province. The top panel plots kernel density estimates implied by the model for selected years. These densities represent the probability distribution from which each year's outcome is drawn, not the empirical cross-section of farm-level yields.

In the 1970s, the estimated time-conditional density is unimodal and tightly concentrated around yields of roughly 2,000 kg/ha (black). By 1979, the entire distribution has shifted right by about 50 percent to center near 3,000 kg/ha (light green) while remaining narrow. By 1989, the distribution has shifted further upward to around 4,000 kg/ha and become notably more dispersed, with an additional concentration of mass near 5,000 kg/ha. This pattern—progressive upward shifts, increasing dispersion, and a persistent notch of mass around the upper quartile—continues through subsequent decades. By 2019 (red), the peak approaches 7,000 kg/ha, and the right tail extends substantially, with non-negligible probability mass beyond 8,000 kg/ha. This evolution reflects both rising central yields and the emergence of a heavier upper tail, indicating that exceptionally high-yield realizations have become increasingly common over time.

The bottom panel overlays the observed annual series (dots), the fitted conditional mean (solid line), and the heteroskedasticity process traced across the distribution. The latter is illustrated with the gray pathways, which are the temporal trajectories of the 500 randomly sampled quantiles that visualize how dispersion evolves at different parts of the distribution. The fan-out is concentrated in the upper quantiles, indicating asymmetric volatility that is stronger in the right tail. Central and lower quantiles display comparatively modest spread, while the fitted mean rises steadily. The

[7] Ongoing work is integrating fine-grained weather, soil, and management data to (i) assess the validity of the constructed counterfactuals, (ii) investigate mechanisms behind heterogeneous treatment effects—such as the roles of irrigation, hybrid selection, and pest pressure—and (iii) refine the donor pool to improve precision and reduce overfitting in the DSC estimates.

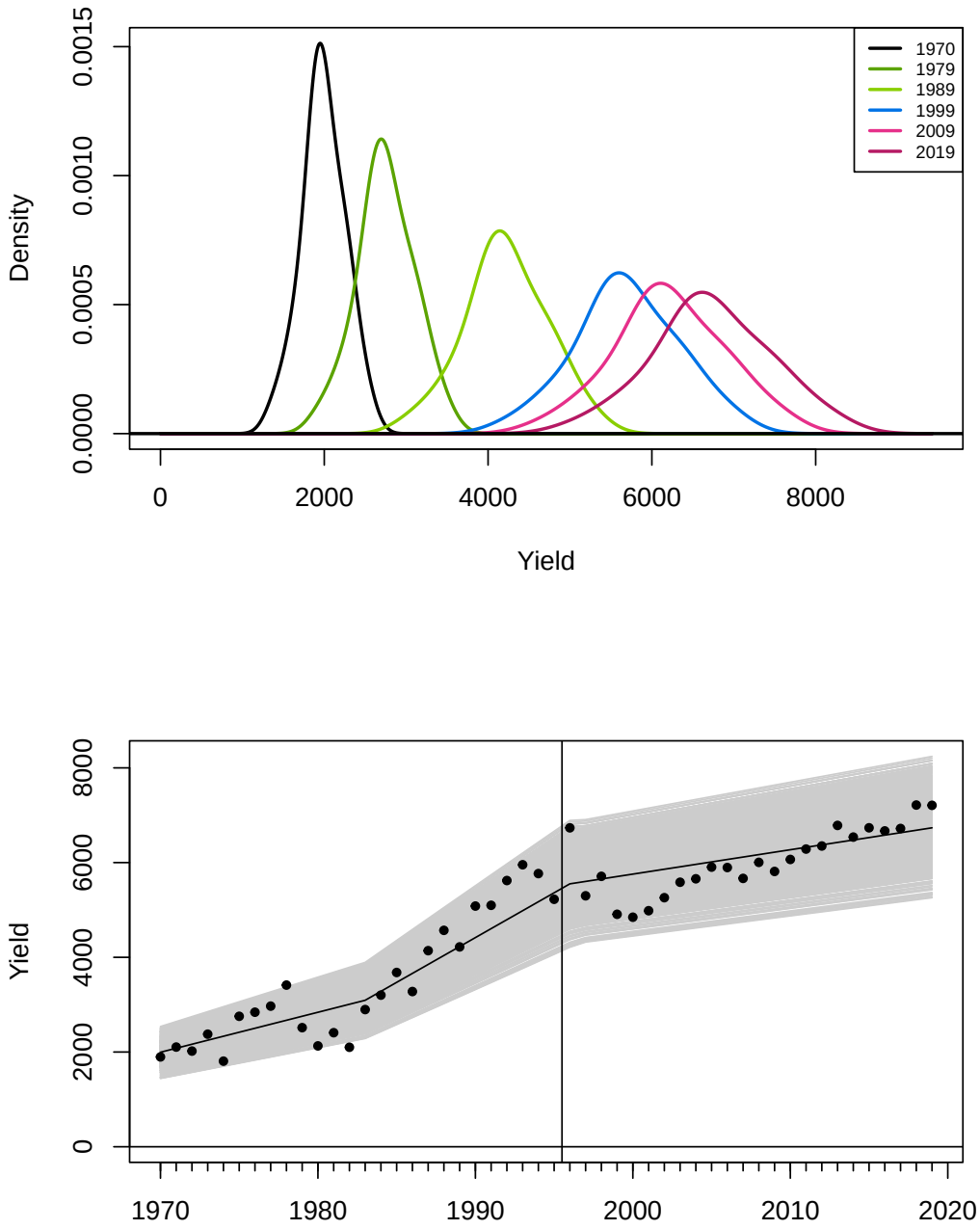


Figure 1: **Estimated time-conditional yield density (top panel) and variance process (bottom panel) for Inner Mongolia.** The top panel shows the model-implied distribution of maize yields for selected years. Each year's yield observation can be interpreted as a draw from its corresponding distribution. The bottom panel plots the estimated trend line and conditional variance. The latter is illustrated with the gray lines, which represent the temporal paths for 500 quantiles randomly drawn from $\mathcal{U}(\varepsilon, 1-\varepsilon)$ with $\varepsilon = 0.025$, illustrating how volatility evolves across the entire yield distribution over time.

pattern is consistent with improvements that moderately favored high-productivity realizations—such as targeted adoption of superior hybrids or management on better-endowed plots—while compressing downside risk.

The patterns are similar for Heilongjiang, Jilin, Shandong, Henan, and Hubei. All exhibit rightward-shifting time-conditional densities and increasing volatility, but the magnitude and timing differ. Heilongjiang's changes are gradual and smooth, Jilin experiences occasional volatility spikes—particularly in the late 1990s and early 2010s—while Shandong and Henan show rapid early gains followed by stabilization. Hubei begins at substantially lower yield levels but records some of the largest relative improvements.

4.2 Counterfactual Outcome Distributions Across Policy Regimes

The DSC estimates (Figures 2–3) suggest that removing the GE ban would have shifted maize yield distributions markedly upward by 2019, with the magnitude and distributional shape of these gains varying considerably across provinces. In several cases—Jilin, Inner Mongolia, Henan, and Hubei—the observed post-treatment distributions lie clearly to the right of their synthetic counterparts, with the largest gaps emerging in the upper tail. This implies that treatment would have increased the probability of exceptionally high yields. In Jilin, gains appear uniform across the support; in the other regions, gains in location are accompanied by gains in dispersion.

The post-treatment patterns rest on a strong foundation of pre-treatment fit. In 1970, treated and synthetic densities are closely aligned in central tendency and tails across all six provinces. By 1996, the match is even stronger, even after the pre-treatment distributions shifting substantially over time. In Hubei the pre-treatment match is less strong, but there remains significant overlap across the two distributions (and places that are more difficult to fit will translate into wider confidence intervals in the re-sampling process). Overall, the close correspondence in the pre-treatment period indicates that the DSC approach successfully reproduces both the level and shape of the yield distribution, so that observed post-treatment divergences are attributable to the policy regime rather than to baseline mismatch.[8]

Looking more closely at the largest producers (Figure 2), Heilongjiang's 2019 distribution is shifted rightward relative to its synthetic, with a modest widening of the upper tail—consistent with

[8]Future work will seek to validate these matches using agro-climatic data.

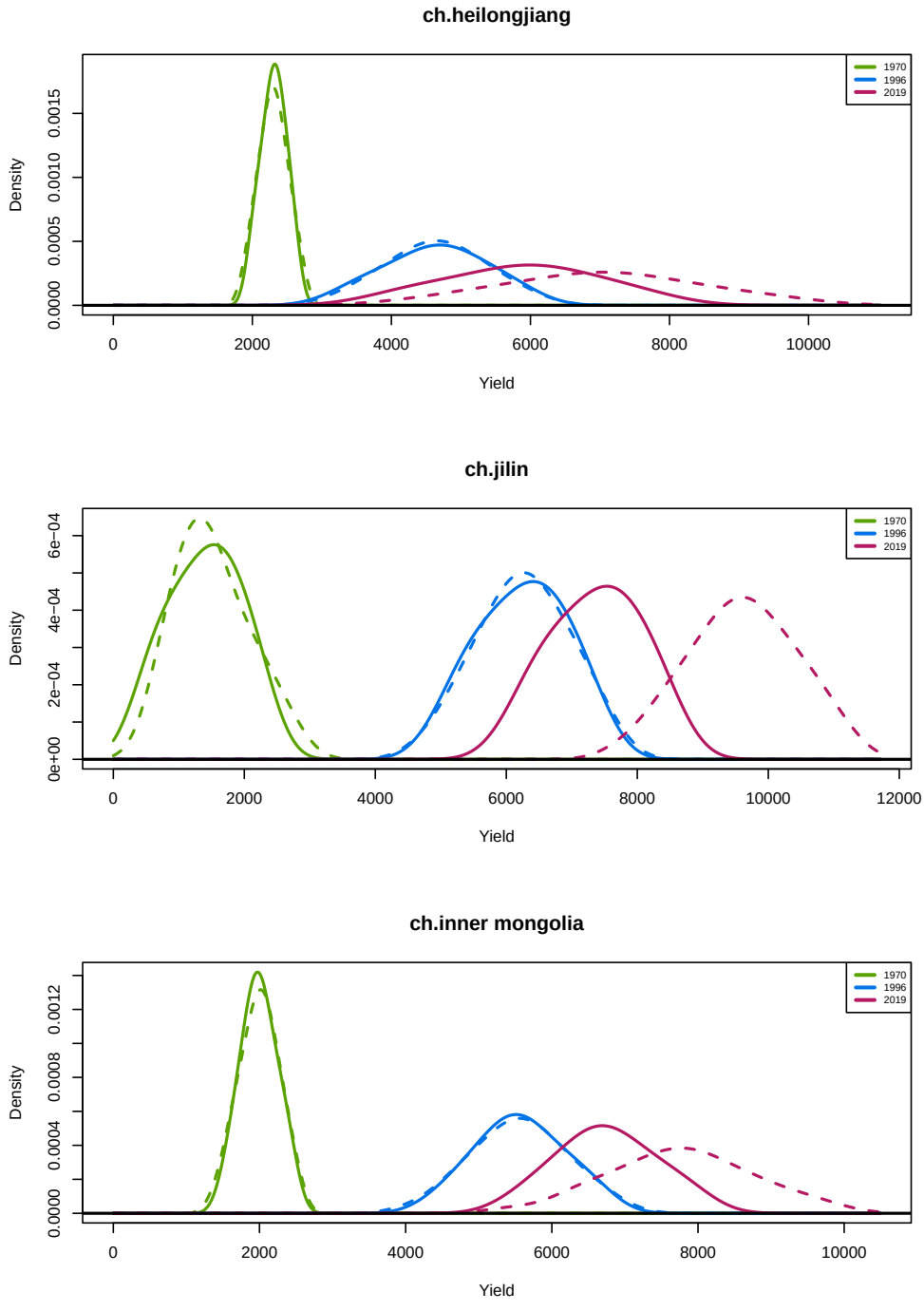


Figure 2: **Distributional synthetic control fits: Heilongjiang, Jilin, and Inner Mongolia.** Treated (solid) and synthetic (dashed) time-conditional densities for 1970 (green), 1996 (blue), and 2019 (red). Pre-treatment fits are close across the support. By 2019, all three provinces show rightward shifts, with Jilin and Inner Mongolia exhibiting pronounced upper-tail gains.

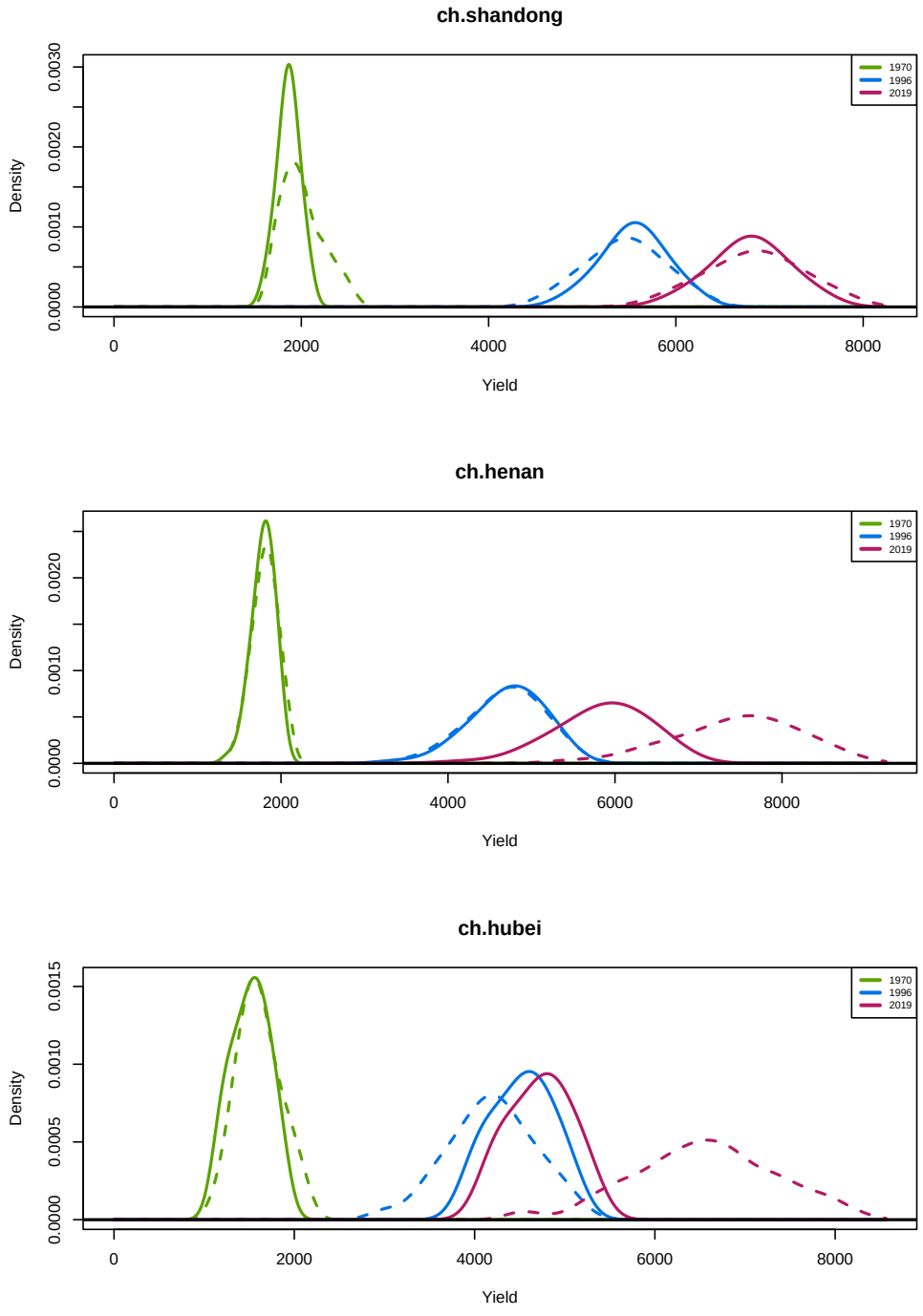


Figure 3: **Distributional synthetic control fits: Shandong, Henan, and Hubei.** Treated (solid) and synthetic (dashed) time-conditional densities for 1970 (green), 1996 (blue), and 2019 (red). Pre-treatment fits are strong. Post-treatment, Shandong has a slight rightward shift in the upper tail, but Henan and Hubei experience broad shifts upwards in location.

a general improvement in yields but only moderate gains in the highest-yielding conditions. Jilin's pattern is more pronounced: the entire distribution shifts rightward and the upper tail becomes substantially heavier than in the counterfactual, indicating a notable rise in the likelihood of extreme yields. Inner Mongolia shows a similarly tail-heavy response, but one that is more concentrated: the lower tail remains closely aligned with its synthetic, while the upper tail stretches well beyond it, suggesting that benefits are concentrated in the most favorable production environments.

For the other provinces (Figure 3), the nature of the shift differs. Shandong's divergence is concentrated in the upper half of the distribution, with the lower half remaining almost unchanged. This pattern points to improvements driven by top-performing plots or conditions, rather than broad-based gains. Henan's distribution, by contrast, shifts rightward across nearly the entire support, implying more uniform productivity improvements that raised both typical and exceptional yields. Hubei stands out: its treated distribution in 2019 is both higher in the center and noticeably narrower in the lower tail. The synthetic retains much more probability mass in low-yield outcomes, suggesting that treatment effects in Hubei reduced the incidence of poor harvests as well as raising average performance.

The province-level patterns visible in Figures 2–3 thus reveal multiple channels through which lifting the GE ban could have affected maize production. In some provinces, especially Inner Mongolia and Jilin, effects are concentrated in the upper tail, implying that technological adoption primarily boosted yields under the best agronomic conditions. In others, like Henan, gains are spread across the distribution, consistent with more general productivity-enhancing effects. And in provinces such as Hubei, the compression of the lower tail suggests an additional risk-reduction channel, lowering the probability of poor outcomes while also improving typical yields.

Taken together, these counterfactual distributions indicate that GE adoption would likely have increased average maize yields by several hundred kilograms per hectare, with the largest proportional improvements in the most favorable growing conditions. At the same time, in provinces like Hubei and Inner Mongolia, the narrowing of the lower tail points to greater yield stability, reducing the downside risk faced by farmers. The DSC framework makes clear that these effects are not limited to a shift in the mean, but involve substantial changes to the entire shape of the yield distribution—effects that would be missed by conventional mean-focused analysis.

The heterogeneous distributional patterns across provinces are consistent with differences in

underlying agronomic and management conditions. In Jilin and Inner Mongolia, the concentration of treatment effects in the upper tail suggests that GE adoption would have generated its largest gains in already favorable environments—possibly through better pest control or improved responsiveness to high input use—while leaving the lower tail largely unchanged. In Henan, broad-based rightward shifts across the distribution point to more uniform productivity enhancements, consistent with technology complementing a wide range of farm conditions. Hubei's combination of higher central yields and lower-tail compression aligns with risk-reducing mechanisms in more marginal environments, such as reduced pest-related yield losses or better stress tolerance under variable weather. While the present analysis does not disentangle these channels, these patterns indicate that both environment-specific advantages and broad-spectrum benefits could be at play, motivating future work that combines DSC with plot-level weather, soil, and management data.

4.3 Changes in Central Tendency

If the GE ban had never been implemented, the estimates in Table 2 suggest that 2019 maize yields in several major Chinese provinces would have been substantially higher. Counterfactual median and mean yields are between 15% and 38% above observed levels in Heilongjiang, Jilin, Inner Mongolia, Henan, and Hubei, with the largest gaps in Hubei ($\approx +38\%$) and Jilin ($\approx +30\%$). Henan (+27.5%) and Heilongjiang (+19%) follow, while Inner Mongolia's difference is more moderate (+15%). In Shandong, the estimated effect is close to zero and statistically insignificant. These counterfactual gains are large when viewed against the cross-country benchmarks in Gollin et al. (2014) and Gollin et al. (2021), where differences of 20–40% in staple grain yields represent a meaningful share of the agricultural productivity gap between upper-middle-income and high-income countries. In that context, removing the ban could have closed a non-trivial share of China's remaining yield gap with the world's most productive maize producers.

The magnitude of our counterfactual effects—15% to nearly 40% increases in median and mean maize yields—are on the upper end of, and in some provinces above, the ranges reported in prior work. Experimental plot evidence from Chavas et al. (2014) suggests GM maize hybrids raise mean yields by single to low double digits, with larger proportional gains in the lower tail. U.S. county-level adoption–intensity estimates in Lusk et al. (2018) imply that moving from zero to full adoption would increase yields by roughly 10–30%, while Scheitrum et al. (2020) find realized gains

Table 2: Effect of GE Ban on Central Tendency of Chinese Maize Yields

	Change in Median (%)		Change in Mean (%)	
Province	Est.	95% CI	Est.	95% CI
Heilongjiang	−18.9	[−29.1, −8.1]	−19.7	[−28.9, −8.5]
Jilin	−29.8	[−39.8, −24.3]	−30.3	[−40.2, −24.7]
Inner Mongolia	−15.2	[−33.5, −12.5]	−14.9	[−32, −11.6]
Shandong	−0.7	[−22.6, 1.6]	−0.7	[−22.1, 1.6]
Henan	−27.5	[−39, −7.4]	−27.5	[−39.4, −7.8]
Hubei	−38.1	[−65.4, −38.2]	−38.6	[−65.8, −37.9]

Notes: Each column reports the point estimate and 95% jackknife confidence interval for a scalar measure of yield distribution change in 2019. Negative values indicate the penalty from ban as a percentage of the scalar measure's baseline value.

of 8–16% in shift–share designs. The multi-country triple-difference estimates in Hansen and Wingender (2023) indicate national-level maize yield gains of about 10–20% from full diffusion, rising in warmer, lower-income, high-pest-pressure environments. Our larger estimates for several provinces likely reflect both agro-ecological suitability and the fact that, unlike shift–share designs, our counterfactual compares full-frontier access to continued zero adoption. In designs where both treated and control regions already have nontrivial adoption, estimated coefficients measure the difference between adoption levels a and b—for example, moving from 60% to 80%—rather than the difference between zero and full adoption. By recovering the yield distribution that would have prevailed absent the 1996 ban, we capture the cumulative effect of decades of trait improvements, hybrid matching, and learning-by-doing, producing larger counterfactual gains than short-run marginal-adoption estimates.

4.4 Changes in Dispersion

Table 3 shows that, in most provinces, the GE ban reduced dispersion in maize yields, with modest effects on the interquartile range (IQR) but large and precisely estimated declines in the coefficient of variation (CV). The IQR effects are generally small—between roughly −1% and −5%—and often statistically indistinguishable from zero. By contrast, CV effects range from −16% to −42% in all provinces except Shandong. Because the CV scales the standard deviation by the mean, these results imply that in the counterfactual without the ban, the standard deviation of yields would have risen proportionally faster than the mean. This does not necessarily indicate an unfavorable change:

Table 3: Effect of GE Ban on Dispersion of Chinese Maize Yields

	Change in IQR (%)		Δ Coefficient of Variation (%)	
Province	Est.	95% CI	Est.	95% CI
Heilongjiang	−1.6	[−10.2, 6.1]	−20.4	[−30.2, −7.9]
Jilin	2.7	[−0.5, 3.2]	−28.8	[−38.9, −23.7]
Inner Mongolia	−3.7	[−7.6, −0.4]	−16.7	[−33.2, −12.8]
Shandong	−3.0	[−5.9, −0.2]	−2.0	[−25.1, −0.1]
Henan	−1.0	[−3.6, 0.6]	−27.9	[−39.2, −8.2]
Hubei	−5.4	[−10.6, 0.2]	−42.3	[−68.9, −40.6]

Notes: Each column reports the point estimate and 95% jackknife confidence interval for the change in a dispersion measure between the observed 2019 yield distribution and the counterfactual distribution without the GE ban. The interquartile range (IQR) change is computed as the proportional change in the distance between the 25th and 75th percentiles, normalized by the median, and the coefficient of variation (CV) change is the proportional change in the ratio of the standard deviation to the mean. Negative values indicate that the ban reduced dispersion relative to the counterfactual.

such a pattern is consistent with broad-based gains accompanied by especially large improvements in the upper tail, as would occur if GE traits deliver their largest benefits under high-potential growing conditions. We next examine changes in the shape of the yield distribution to distinguish between upward stretching driven by positive tail gains and increases in downside risk.

4.5 Changes in Shape

From an economic perspective, the aspects of the yield distribution most relevant to farmers and policymakers are those that shape production risk—how severe poor outcomes are, how stable typical yields remain, and how much scope there is for exceptional harvests. These dimensions matter directly for farm income volatility, crop insurance payouts, and the profitability of risk-mitigation strategies.

Table 4 and Figure 4 reveal that the GE ban altered the shape of the maize yield distribution in ways that are economically meaningful and geographically heterogeneous. Starting with the 20th percentile, which captures performance in relatively poor years, the effects are uniformly negative in five of the six provinces. The largest losses occur in Hubei (−33.6%) and Jilin (−32.8%), where the ban lowered this lower-tail benchmark by roughly one-third. Heilongjiang (−18.7%) and Henan (−26.2%) also show large and statistically significant declines, while Inner Mongolia's loss is somewhat smaller at −11.6%. Only Shandong records a small and statistically insignificant change (+1.4%), consistent with its muted response in previous tests. These results indicate that GE

Table 4: Changes in Yield Distribution Shape Under GE Ban

	Δ 20th Percentile (%)		Δ 80th Percentile (%)		Δ Prob(Low yield) (pp)	
Province	Est.	95% CI	Est.	95% CI	Est.	95% CI
Heilongjiang	−18.7	[−31.1, −8.2]	−20.8	[−31, −7.4]	17.5	[9.4, 24.4]
Jilin	−32.8	[−43.8, −26.3]	−28.9	[−39.1, −23.7]	16.1	[12, 27.9]
Inner Mongolia	−11.6	[−29.9, −9.2]	−17	[−32.9, −12.8]	11.1	[3.5, 18]
Shandong	1.4	[−18.4, 4.4]	−2.3	[−25.6, −0.6]	−2.7	[−10.7, 10.1]
Henan	−26.2	[−38.9, −6.5]	−28	[−39.4, −8.4]	12.6	[5.7, 15.3]
Hubei	−33.6	[−61.7, −28.1]	−43.6	[−70.2, −41.3]	10.5	[6.3, 14.7]

Each column reports the point estimate and 95% jackknife confidence interval for the difference between the observed 2019 yield distribution under the GE ban and the counterfactual without the ban. The 20th and 80th percentile changes are proportional differences in yields at those quantiles; negative values indicate that the ban reduced yields at that point in the distribution. "Prob(low yield)" is the change (in percentage points) in the probability of producing less than 90% of the mean yield, a threshold analogous to a crop-insurance trigger; positive values indicate increased downside risk under the ban.

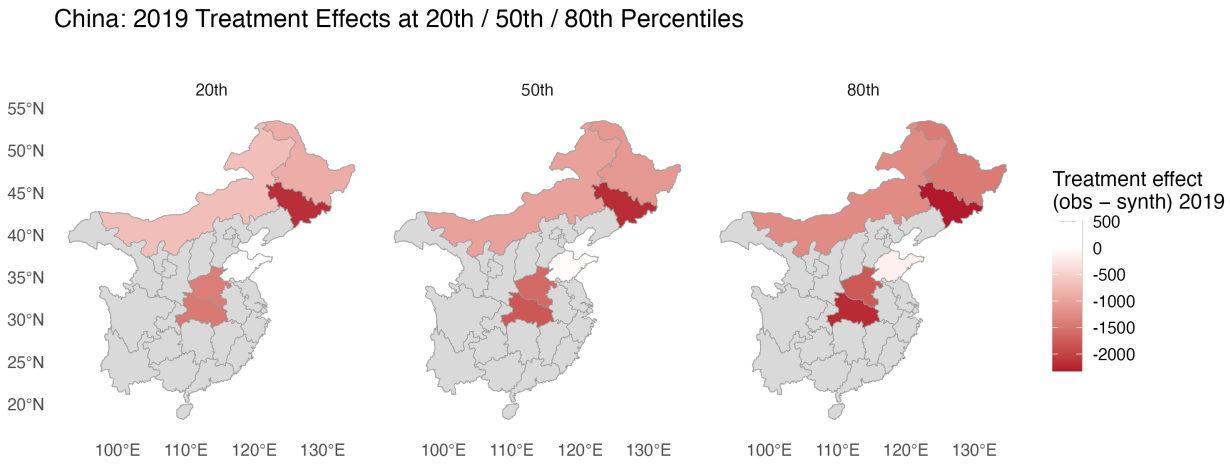


Figure 4: **Map of estimated foregone annual yields in 2019 across provinces.** The difference in yields between the observed distributions and the synthetic control distribution evaluated at the 20th, 50th, and 80th percentiles. Values are in kilograms per hectare. Darker red shading indicates larger yield shortfalls under the ban at that percentile. Negative values across all three percentiles indicate that the ban simultaneously lowered the floor, median, and ceiling of potential yields, with the largest proportional losses occurring in the upper tail in most provinces.

access would have substantially raised the floor on yields in bad years for most major producing provinces.[9]

Turning to the 80th percentile, which reflects outcomes in good years, the losses under the ban are even larger in most provinces. Hubei again leads with a -43.6% shortfall, meaning that its high-yield years would have been nearly half again as productive absent the ban. Jilin and Henan follow closely at -28.9% and -28.0%, respectively, while Heilongjiang's loss is -20.8%. Inner Mongolia sees a -17.0% decline, and even Shandong—largely unaffected at the 20th percentile—shows a statistically significant -2.3% loss at the upper tail. This pattern suggests that much of the increase in dispersion documented earlier reflects favorable "stretching" of the distribution driven by especially strong gains in high-potential conditions.

The probability of a low yield (below $0.9\times$mean) rises significantly under the ban in every province except Shandong, by margins of 10–17 percentage points. In Hubei, the ban raises this probability by over 10 points, while in Heilongjiang and Jilin, the increases are 17.5 and 16.1 points, respectively. Inner Mongolia (+11.1) and Henan (+12.6) also see sizable rises in the risk of falling below the crop-insurance-style loss threshold. The fact that the ban simultaneously reduced high-percentile yields and raised the likelihood of triggering a loss payment underscores its broad risk-increasing effect.

Figure 4 visually reinforces these findings. The 20th-percentile map shows darker shading in Jilin, Hubei, and Heilongjiang, highlighting the provinces where the ban most depressed the lower tail. The 50th-percentile panel, the median effect, shows the earlier central-tendency results, with Hubei, Jilin, and Henan showing the deepest losses. The 80th-percentile map is darker holding province constant, confirming that the ban's largest proportional effects occur in favorable years. This combination—raising the risk of poor outcomes while capping the best outcomes—is exactly the risk profile farmers and policymakers would prefer to avoid.

In sum, the shape-analysis results show that the GE ban did more than shift the average yield downwards: it altered the structure of production risk in a way that is unfavorable to both farmers

[9]We also tested for overall differences between the observed and counterfactual yield distributions using distributional change tests. Such tests are inherently low power, since inference must account for variation across the full support, and even economically meaningful divergence can fail to register as statistically significant. Consistent with this, the average W2 and quantile-by-quantile tests detected significant differences only for Jilin. Other provinces show post-ban divergence from their counterfactuals—often concentrated in higher-yield quantiles—but magnitudes remain below conventional significance thresholds. Future work will reconsider these tests under more powerful alternatives for distributional differences, such as supremum-type tests.

and insurers. Without the ban, provinces would have seen both a higher floor in bad years and a higher ceiling in good years, lowering the frequency of insurance-triggering shortfalls while allowing more frequent bumper harvests.

4.6 Summary

Three main findings emerge. First, time-conditional density estimates show a long-run pattern of rising central yields in all six provinces, with heavier upper tails and, in some cases, reduced lower-tail mass. Second, counterfactual simulations indicate that lifting the GE ban would have shifted the entire yield distribution upward by 2019, with some provinces—such as Jilin and Inner Mongolia—seeing the largest gains in the upper tail, and others—such as Henan and Hubei—experiencing more uniform improvements across the distribution. Third, scalar tests of central tendency, dispersion, and shape confirm that these shifts are statistically and economically important: absent the ban, median and mean yields would be 15–38% higher in most provinces, the probability of low yields would be substantially reduced, and the best-year outcomes would be markedly better.

These counterfactual effects are at the high end of, and in some cases exceed, yield gains reported in the existing GE maize literature. Experimental trial results such as Chavas et al. (2014) typically find mean gains in the single to low double digits, while U.S. adoption-intensity designs (Lusk et al., 2018) imply 10–30% increases when extrapolated from zero to full adoption, and Scheitrum et al. (2020) report realized gains of 8–16% in shift–share designs. Multi-country triple-difference estimates from Hansen and Wingender (2023) suggest national-level maize yield gains of 10–20% from full diffusion, with larger effects in high-pest-pressure environments. Our larger estimates likely reflect both the agro-ecological suitability of major maize-producing provinces and the fact that our design compares full-frontier access to sustained zero adoption, capturing the cumulative effect of decades of trait improvements and hybrid adaptation rather than marginal short-run changes.

5 Conclusion

Long-term restrictions on agricultural biotechnology alter not just average productivity, but also the entire distribution of outcomes, shaping both risk and stability in meaningful ways. We quantify

the effects of China's long-standing ban on GE maize using a distributional synthetic control framework applied to province-level yield data from 1970–2019. By matching the full pre-treatment yield distributions of six major maize-producing provinces to a weighted combination of GE-adopting regions in the United States, Argentina, and Brazil, we recover counterfactual distributions for how yields would have evolved absent the ban. These counterfactuals abstract from political economy, consumer acceptance, and export-market certification constraints; our estimates isolate the technological consequences conditional on adoption.

By recovering the full dynamic trajectory of the counterfactual yield distribution, the analysis moves beyond conventional mean-focused approaches to capture effects on central tendency, dispersion, and tail probabilities that matter for both productivity and risk. The results indicate sizable and heterogeneous impacts. In most provinces, median yields would have been substantially higher without the ban, with even larger proportional gains in the upper tail. In five of the six provinces, the lower tail is shifted inward, indicating a reduction in downside risk. These effects imply not only foregone efficiency gains but also foregone improvements in yield stability, both central to food security policy.

The relevance of these findings extends beyond China. Many countries, including Peru, Venezuela, Ecuador, India, Kenya, and Zambia, maintain full or partial bans on GE cultivation despite similar pressures to raise productivity and manage production risk. Our estimates suggest that such policies can be associated with large opportunity costs: they may reduce not only average output but also the likelihood of exceptional harvests, while increasing the probability and severity of poor outcomes. In contexts where risk management and stability are valued alongside output growth, these distributional effects are at least as important as changes in the mean.

Methodologically, the paper adapts DSC to settings with one observation per unit-year and incorporates a clustered jackknife to propagate uncertainty from the estimation of conditional quantile functions into all downstream statistics. It also develops a pooled, time-invariant donor-weight estimator that solves an open problem left by Gunsilius (2023) on stabilizing weights over long pre-treatment periods. This reduces overfitting and improves interpretability. These innovations are likely to be useful for policy evaluation in other agricultural and non-agricultural contexts where full distributions, rather than just means, are of interest.

We are currently extending this analysis in several directions. First, we are refining the donor

pool by adding high-quality non-adopting but high-productivity regions and trimming donors that contribute to overfitting. We are also implementing robustness exercises such as nearest-neighbor donor restrictions, country-exclusion tests, and placebo re-timings to sharpen the assessment of the external validity and stability of the DSC estimates. Second, we are incorporating fine-grained weather, soil, and management data to investigate mechanisms behind the heterogeneous distributional effects we document, including whether upper-tail gains are linked to irrigation, hybrid choice, or pest pressure, and whether lower-tail compression arises in more marginal environments. Third, we are conducting additional diagnostic exercises, including in-time placebos, leave-one-donor-out tests, alternative distance measures, and formal balance tables comparing treated and synthetic units, to further assess the credibility of the counterfactual construction.

Restricting GE maize in China has been associated with substantial losses in both average yields and yield stability, with gaps that would be expected to widen if the ban had persisted. The findings illustrate that biotechnology policy can shape the entire distribution of agricultural outcomes, and that analyses focusing only on mean effects may understate the full costs of foregone technological adoption.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132.

Alston, J. M. and Pardey, P. G. (2014). Agriculture in the global economy. *Journal of Economic Perspectives*, 28(1):121–146.

Areal, F., Riesgo, L., and Rodriguez-Cerezo, E. (2013). Economic and agronomic impact of commercialized GM crops: a meta-analysis. *Journal of Agricultural Science*, 151(1):7–33.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012.

Chavas, J.-P., Shi, G., and Lauer, J. (2014). The effects of GM technology on maize yield. *Crop Science*, 54(4):1331–1335.

Chen, J. (2023). Synthetic control as online linear regression. *Econometrica*, 91(2):465–491.

de Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Evenson, R. E. and Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science*, 300(5620):758–762.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Gollin, D., Hansen, C. W., and Wingender, A. M. (2021). Two blades of grass: The impact of the green revolution. *Journal of Political Economy*, 129(8):2344–2384.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The agricultural productivity gap. *Quarterly Journal of Economics*, 129(2):939–993.

Gunsilius, F. F. (2023). Distributional synthetic controls. *Econometrica*, 91(3):1105–1117.

Hansen, B. E. (2022). Jackknife standard errors for clustered regression. Technical report, Working paper, University of Wisconsin.

Hansen, C. W. and Wingender, A. M. (2023). National and global impacts of genetically modified crops. *American Economic Review: Insights*, 5(2):224–240.

Harri, A., Coble, K. H., Ker, A. P., and Goodwin, B. J. (2011). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, 93(3):707–717.

Ker, A. P. and Tolhurst, T. N. (2019). On the treatment of heteroscedasticity in crop yield data. *American Journal of Agricultural Economics*, 101(4):1247–1261.

Lusk, J. L., Tack, J., and Hendricks, N. P. (2018). Heterogeneous yield impacts from adoption of genetically engineered corn and the importance of controlling for weather. In *Agricultural productivity and producer behavior*, pages 11–39. University of Chicago Press.

Qaim, M. and Zilberman, D. (2003). Yield effects of genetically modified crops in developing countries. *Science*, 299(5608):900–902.

Scheitrum, D., Schaefer, K. A., and Nes, K. (2020). Realized and potential global production effects from genetic engineering. *Food Policy*, 93:101882.