The effect of conservation tillage on corn and soybean yields in the US Corn Belt: a Post-Double-Selection method

Menglin Liu

Ph.D. Student

Department of Agricultural and Consumer Economics

University of Illinois at Urbana-Champaign

mliu11@illinois.edu

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Abstract

Global warming impacts everyone, private initiatives and the government are taking actions to prompt regenerative agricultural practices that can sequester carbon in the soil and vegetation to tackle climate change. It is critical to understand the yield effects of these practices on crops to facilitate the promotion. Many findings on yield effect of conservation tillage (CT) are based on the field experiments, only a few observational studies have been done. This study uses county-level data covering 631 counties across 12 states from 2005 to 2018 and the post-double-selection method to examine the average effect of CT on corn and soybean yields. I find that CT has no impact on corn but small negative impact on soybeans. In counties with an average CT adoption rate, the soybean yields would decline by 3.2%, this can be translated to an average loss of 1.5 bushels per acre. The average loss for each county due to 10% increase in CT adoption for soybeans could be 346,720 US dollars based on the latest price data.

Keywords: Conservation Tillage, Yield, Corn, Soybeans, Post-Double-Selection

1 Introduction

Global warming is mainly driven by emissions from human activities (Masson-Delmotte et al., 2021). Climate change caused by global warming has increased the frequency and intensity of weather and climate extremes such as heavy precipitation, droughts, and hot extremes (Masson-Delmotte et al., 2021). These climate hazards have negative impacts on human's life, and increase the risk of food security; thus, actions are needed to slow down the climate change induced by human activities. As part of the efforts to tackle climate change, private initiatives and the government are seeking ways to reduce Greenhouse Gases emissions and enhance the agricultural carbon sink. In recent years, the private sector has initiated several carbon emissions offset and supply chain inset markets (Bruner & Brokish, 2021; Plastina & Wongpiyabovorn, 2021; Thompson et al., 2021) to incentivize farmers to adopt practices such as conservation tillage and cover crops that sequester atmospheric carbon in soils and vegetation. The US president's Climate 21 Project also called for the establishment of a 'carbon bank' that would pay farmers to store carbon using regenerative agriculture practices. Conservation tillage (CT), which leaves at least thirty percent of the soil surface covered by residue after planting (CTIC, 1992), is one such practice that has been recommended for adoption for years because of its soil conservation benefits. However, CT may reduce yields due to weed and insect problems, low soil temperature, and high soil moisture caused by residues (Toliver et al., 2012), thus, farmer concerns about yield loss following adoption have been a major impediment to expanding adoption of this practice (Kragt et al., 2017; Prokopy et al., 2019). Understanding the effects of CT adoption on corn and soybean yields is necessary to quantify the economic impact of CT adoption on farmers, and evaluate carbon market prices or conservation payments to farmers for adopting CT.

Many studies have been done based on field experiments to assess the yield effects of CT, however, the results are mixed because yield effects can be influenced by other factors, yields for crops with CT tend to be higher under crop rotation, with well-drained soil, and in warmer locations (DeFelice et al., 2006; Toliver et al., 2012). Though side-by-side field experiments provide direct assessment of the yield effects of CT, the results of these studies can have limited real-world implications because same management for CT and conventional tillage was used in many field experiments (DeFelice et al., 2006), and the field experiments were conducted in certain locations (Chen et al., 2021), therefore, the results from these experiments cannot be generalized. Large-scale observational data that captures real-world applications in different locations may be used to address these limitations and identify the average effect of CT adoption on corn and soybean yields.

There are only a few observational studies on this topic, two most recent ones are Deines et al. (2019) and Chen et al. (2021). Deines et al. (2019) used the fine-scale satellite imagery data of tillage practices and model derived crop yields in the US Corn Belt from 2005 to 2017, and applied casual forest method to assess the yield effects of CT. They found that CT has slightly positive yield effect for corn and insignificant effect for soybeans in the short run. (Chen et al., 2021) used the county-level tillage data from the Operational Tillage Information System (OpTIS) in the US Corn Belt and yield data from USDA survey from 2005 to 2018, and specified a fixed-effects linear model to estimate the effects of CT adoption rates on corn and soybean yields. They found that CT has no negative yield effects for corn or soybeans. Both studies have limitations, first, they included no or only a few state-level and national level socioeconomic variables that influence CT adoption and yields in their analysis, this could cause omitted-variables bias (Chen et al., 2021; Deines et al., 2019), many studies have shown that CT adoption and crop yields can be influenced by socioeconomic variables (Cabas et al., 2010; Kaufmann & Snell, 1997; Prokopy et al., 2019); second, the linear model used in Chen et al. (2021) might not be the correct specification for this problem, previous studies including the quadratic terms of weather and socioeconomic variables to predict crop yields found significant influences of quadratic terms on yields (Cabas et al., 2010; Tannura et al., 2008). This work contributes to these existing observational studies on yield effects of conservation tillage adoption using a county-level socioeconomic dataset and applying new empirical methods that allow for nonlinear trends interacted with observed county-level time varying characteristics. This study applies post-double-selection method (Belloni et al., 2014a) to a high-dimensional regression model to examine the effects of CT adoption on county mean corn and soybean yields by using corn and soybean yields, CT adoption rates, weather and climate levels, and socioeconomic data from 631 counties across twelve Corn Belt states from 2005 to 2018. I apply LASSO to the model for variable selection and predicting yields and CT adoption rates, respectively, and identify the effect of CT adoption on corn and soybean yields by regressing the differences of yields on the differences of CT adoption rates and the union of the selected variables and the year fixed effect dummies.

This study suggests that CT has no impact on corn but small negative impact on soybeans. In counties with an average CT adoption rate, the soybean yields would decline by 3.198% (CI = [4.577%, 1.820%]), this is equivalent to an average loss of 1.518 bushels per acre (CI = [2.173, 0.864]). Farmers adopt CT might be because of the benefits from cost reduction through using less fuel, labor, and insecticide; and the incentives from government payments. Smaller confidence intervals and RMSE are obtained when using post-double-selection method, this suggests that double selection method could provide more accurate estimations.

I discuss data in the next section, followed by methods used, results, discussions and conclusions.

2 Data

This study examines the yield effects of CT adoption rates on corn and soybeans by using: crop yields from USDA-NASS (2020b); CT adoption rates from the Operational Tillage Information System (OpTIS) (OpTIS, 2019); weather and climate data from the PRISM database (PRISM Climate Group) and TerraClimate (Climatology Lab); and socioeconomic data obtained from U.S. Bureau of Economic Analysis (BEA) (2021), USDA-Risk Management Agency (USDA-RMA) and USDA-NASS (2020a). I use some data used in Chen et al. (2021) such as corn and soybean yields, CT adoption rates, standardized precipitation-evapotranspiration index (SPEI), price received, continuous crop (corn/soybeans), fraction insured, and GM adoption rates. In addition, monthly weather and derived climate levels, Growing Degree Days (GDD), farm incomes, farm expenses, and values of inventory changes are used in this analysis. The data cover 631 counties across twelve Corn Belt states (Iowa, Illinois, Indiana, Kansas, Michigan, Minnesota, Missouri, Nebraska, Ohio, Oklahoma, South Dakota, Wisconsin) from 2005 to 2018. In the following subsections, I discuss the dependent and independent variables, and explanatory variables such as weather and climate variables and socioeconomic variables that are included in this analysis, and the limitations of the data.

2.1 Dependent and independent variables

I use county mean corn and soybean yields as the dependent variables, and county-level corn and county-level soybean CT adoption rates as the independent variables, respectively, for corn and soybean regressions. The yields are from the survey data published on USDA-NASS; the corn yields and soybean yields across counties and years vary from 19 bushel/acre to 246.7 bushel/acre and from 10.9 bushel/acre to 82.3 bushel/acre, respectively. CT adoption rates are obtained by aggregating remotely sensed corn and soybean tillage data from crop level to county level, the CT adoption rates for corn and soybean vary from 0.01% to 100% and from 0.01% to 100%, respectively. The summary statistics of yields and CT adoption rates can be found in Appendix Table A2.

2.2 Explanatory variables

2.2.1 Weather and climate variables

Temperature and precipitation have been used in many studies to predict crop yields; some studies include both seasonal climate variables and monthly climate variables (Cabas et al., 2010; Deines et al., 2019; Tannura et al., 2008), and some studies use aggregated climate proxies such as Growing Degree Days (GDD) during growing season (Anandhi, 2016; Deines et al., 2019), while some studies also include climate variables out of crop growing season (Dixon & Segerson, 1999; Tannura et al., 2008). To capture the potential effects of different forms of temperature and precipitation variables, I obtain the monthly weighted average precipitation, minimum and maximum temperature; and Growing Degree Days (GDD) during growing season from April to October calculated based on the daily precipitation, minimum temperature, and maximum temperature from the original data sources.

In addition to temperature and precipitation, solar radiation is found to be an important factor in determining crop yields (Dixon, 1994; Lobell et al., 2009). Precipitation, as one of the commonly used factors to predict crop yields, may not represent the water condition for crop production since water stress can be induced by increased vapor pressure deficit (VPD) (Zhang et al., 2017) and water deficits in rainfed system could increase the yield gap (Lobell et al., 2009). Palmer Drought Severity Index (PDSI) is a common variable used to measure drought, and it is highly correlated with soil moisture (Mika et al., 2005); early season soil moisture (April to June) are found to have significant effects on corn and soybean yields (Urban et al., 2015). While Lobell et al. (2014) found that VPD is a better predictor of crop water stress and yields than PDSI. Farmers may adopt CT to reduce the risk of yield loss due to drought conditions because Chen et al. (2021) found that CT adoption may mitigate some of the negative impacts of drought on corn and soybean yields. Runoff as a measure of both waterflow to other areas and degradation of top soil through leaching and erosion (Kerr, 2007) may influence both crop yields and CT adoption since floods can reduce yields and farmers are more likely to adopt CT on high soil erosion land (Wu et al., 2004). Wet soils in fall after harvesting and in Spring before planting may prevent farmers from performing tillage (Chen et al., 2021), thus, these conditions may have influence on CT adoption.

Climate Water Deficit (DEF), soil moisture, PDSI, VPD and runoff are derived directly or indirectly based on the combinations of some of the primary climate variables including maximum temperature, minimum temperature, vapor pressure (VAP), precipitation accumulation, downward surface shortwave radiation (SRAD), and wind-speed (WS), respectively. Reference evapotranspiration (PET) and actual evapotranspiration (AET) are two factors that determine DEF, they are also derived based on the primary climate variables (Abatzoglou et al., 2018). Obviously, the primary climate variables are correlated with one or more than one derived climate variables. However, it is hard to determine if the primary climate variables are better predictors of yields or if the one-year lag of the derived climate variables such as drought conditions, soil moisture, and runoff can be better perceived by farmers to influence their CT adoption decisions, and vice versa. In order to not miss the important information that influence yields or CT adoption, I include all these variables in the study and rely on the variable selection method to select the variables that are most important to predict yields or CT adoption.

The weather variable county-level means, such as monthly precipitation, monthly temperature, and GDD for the growing season, are averaged or accumulated daily weighted means calculated using R package ‘acdcR’ (Yun, 2022). The derived climate variable county-level means are monthly weighted means calculated following Yun and Gramig (2019). Both weighed calculations are only accounted for the agricultural fields and are calculated using the agricultural field areas in the PRISM or TerraClimate data pixels as weight in the county.

2.2.2 Socioeconomic and crop management variables

Previous studies found that crop insurance, crop rotation, income, and input costs may be correlated with CT adoption (Prokopy et al., 2019). High fuel costs and labor costs may incentivize farmers to choose lower intensity tillage practices (Chen et al., 2021; Deines et al., 2019; Uri, 2000), however, higher fertilizer prices may discourage farmers to adopt CT since farmers tend to use more fertilizer in conservation tillage system (Laukkanen & Nauges, 2011). Studies have shown that crop rotation adoption has influence on yields (Al-Kaisi et al., 2015; Karlen et al., 2013; West et al., 1996) and CT adoption (Wu & Babcock, 1998). Input costs such as seed expenses, fertilizer and lime expenses, petroleum product expenses, machinery expenses, other expenses and value of materials and supplies inventory change can be treated as proxies for the input usage which have a significant impact on corn and soybean yields (Van Roekel et al., 2015). Technology adoption such as using Genetically Modified (GM) seeds could influence crop yields (Lusk et al., 2017) and the adoption of CT (Chen et al., 2021); farm’s financial position such as the lags of incomes and value of inventory changes may represent the farm’s technology adoption such as adoption of precision agriculture and cropping management capabilities which could also have impacts on crop yields (Egli, 2008). Thus, I include socioeconomic and crop management variables such as the lags of farm incomes and value of inventory changes, farm expenses, corn and soybean insured ratios, GM adoption, and crop rotation proxies (continuous corn to corn and continuous soybean to soybean ratios) in the study, details can be found in Appendix table 1.

2.3 Data limitations

County-level aggregated data cannot provide farm-level characteristics information, such as farmer's age and education level, and farm-level management information. I assume that farmer's personal characteristics are controlled since they may not be time-varying characteristics, or they may change in a very small pace each year that can be ignored. Farmer's awareness of environmental issues, perceptions on climate change, and attitudes toward conservation practices are very hard to be captured, however, these may affect the farmer's decision on CT adoption.

There are measurement errors in the CT adoption rates, these errors are introduced by using the rates calculated based on the observed residue cover in part of the county to replace the missing observations for the reminder of the county (Chen et al., 2021). To address this limitation, I conduct robustness checks by excluding states with high missing data rates.

3 Methods

3.1 Methodology

As discussed earlier in the data section, many factors may be associated with yields and CT adoption rates, thus, I include as many weather and socioeconomic variables that might be relevant as possible in this analysis. To capture the evolution of yields and CT adoption rates over a 14-year period, I adopt a model that allows for nonlinear trends interacted with observed county-specific time-varying characteristics (Belloni et al., 2014b). High-dimensional data are obtained by interacting and transforming the variables. To avoid collinearity and overfitting, simply dropping some variables based on economic intuition or using a linear model to form a small number of interaction terms might be problematic since it is hard to determine if correct variables and functional forms were chosen. High-dimensional methods can use the data to learn which variables are most important (Belloni et al., 2014b) and pick one variable among very correlated predictors (Friedman et al., 2010); this helps us select the variables that are most useful for predicting yields and CT adoption rates, respectively, and avoid collinearity and overfitting. Because certain factors may be associated with both yields and CT adoption rates, a naive approach to apply variable selection methods to a single regression model may lead to omitted-variable bias if any variable that is highly related to CT adoption rates is dropped (Belloni et al., 2014b) and lead to regularization biases because of variable selection (Chernozhukov et al., 2018). Thus, I choose the double selection methods to overcome these biases.

I apply post-double-selection method (Belloni et al., 2014a) to a high-dimensional regression model (Belloni et al., 2014b) following the concept of partially linear regression (PLR) model (Belloni et al., 2014a; Robinson, 1988) to examine the yield effect of CT adoption rates on county mean corn and soybean yields in the Corn Belt.

$$\Delta y_{it} = \beta_0 + \alpha \Delta CT_{it} + \boldsymbol{z}'_{it}\boldsymbol{\beta} + \mu_t + \Delta\varepsilon_{it}, \tag{1}$$

$$\Delta CT_{it} = \gamma_0 + \boldsymbol{z}'_{it}\boldsymbol{\gamma} + \nu_t + \Delta\epsilon_{it}, \tag{2}$$

Where Δy_{it} = y_{it} - y_{it-1}, y_{it} is the natural log of crop (corn or soybean) yield in county i in year t; β_0 and γ_0 are constant terms; ΔCT_{it} = CT_{it} - CT_{it-1}, CT_{it} is the CT adoption rate in i and t; α is a coefficient to be estimated; $\boldsymbol{z}'_{it}\boldsymbol{\beta}$ and $\boldsymbol{z}'_{it}\boldsymbol{\gamma}$ are approximations to the true functions which are unknown, $\boldsymbol{z}_{it}$ is the set of the technical regressors[1] that includes variables made up of the differences and lags of weather and climate levels and values of socioeconomic factors, initial levels (the level in the first year) and initial differences (the difference of the levels between the second year and the first year) of weather and socioeconomic variables, the initial level and initial difference of the CT adoption rate, quadratics in each of the preceding variables, interaction of all differences, and interactions of all the variables with t and t^2. This set of technical regressors corresponds to a cubic trend for the level of crop yield and CT adoption rate that is allowed to depend on the observed county-level characteristics. μ_t and ν_t are time fixed effects. $\Delta\varepsilon_{it}$ and $\Delta\epsilon_{it}$ are error terms which are defined similarly to Δy_{it} and ΔCT_{it}.

To apply the variable selection method, I write the reduced form corresponding to (1) as:

$$\Delta y_{it} = \bar{\beta}_0 + \boldsymbol{z}'_{it}\bar{\boldsymbol{\beta}} + \bar{\mu}_t + \Delta\bar{\varepsilon}_{it}, \tag{3}$$

where $\bar{\beta}_0 = \alpha\gamma_0 + \beta_0$, $\bar{\boldsymbol{\beta}} = \alpha\boldsymbol{\gamma} + \boldsymbol{\beta}$, $\bar{\mu}_t = \alpha\nu_t + \mu_t$, $\Delta\bar{\varepsilon}_{it} = \alpha\Delta\epsilon_{it} + \Delta\varepsilon_{it}$.

I assume approximate sparsity condition hold for this case, which means that only a relatively small number of non-zero coefficients in $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ could make the approximation errors small relative to estimation error. This assumption may enable ΔCT_{it} be taken as exogenous conditional on a relatively small number of control variables and ensure that important elements are captured by a relatively small number of control variables to keep the residual variance small when regressing Δy_{it} on ΔCT_{it}. The Least Absolute Shrinkage and Selection Operator (LASSO) is used to select a relatively small and approximately right set of technical regressors to predict yields and CT adoption rates, respectively. I apply LASSO to equations (2) and (3) to select

[1] Technical regressors refer to the regressors obtained through transformations and interactions of the control variables, which is recommended by Belloni et al. (2014a).

technical regressors that are useful for predicting ΔCT_{it} and Δy_{it}. The LASSO estimators $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ (Friedman et al., 2010) are solutions for the following optimization problems, respectively:

$$\min_{(\gamma_0,\boldsymbol{\gamma})\in \mathrm{R}^{p+1}} \left[\frac{1}{2(T\times I)}\sum_{t=1}^{T}\sum_{i=1}^{I}(\Delta CT_{it}-\gamma_0-\boldsymbol{z}_{it}'\boldsymbol{\gamma}-\nu_t)^2+\lambda_\gamma\sum\nolimits_{j=1}^{p}\left|\gamma_j\right|\right], \tag{4}$$

$$\min_{(\bar{\beta}_0,\bar{\boldsymbol{\beta}})\in \mathrm{R}^{p+1}} \left[\frac{1}{2(T\times I)}\sum_{t=1}^{T}\sum_{i=1}^{I}(\Delta y_{it}-\bar{\beta}_0-\boldsymbol{z}_{it}'\bar{\boldsymbol{\beta}}-\bar{\mu}_t)^2+\lambda_{\bar{\beta}}\sum\nolimits_{j=1}^{p}\left|\bar{\beta}_j\right|\right], \tag{5}$$

where λ_γ and $\lambda_{\bar{\beta}}$ are the "penalty level", I is the total number of counties, T is the total number of years, and p is the total number of covariates, γ_0 and $\bar{\beta}_0$ are constant terms. LASSO performs variable selection by forcing the coefficients of "not-so-significant" variables to become zero through the penalty. After the important regressors are selected using LASSO for predicting ΔCT_{it} and Δy_{it}, I use the union of the set of selected technical regressors, including time fixed effects, as controls to estimate α in the ordinary least squares regression of Δy_{it} on ΔCT_{it}. This approach ensures that any technical regressors that have large effects on Δy_{it} and ΔCT_{it} are included in the model and omitted-variable bias is limited by excluding technical regressors that are most mildly associated with Δy_{it} and ΔCT_{it}.

To test the performance of this post-double-selection approach, I also specify a model with first differenced and lagged control variables for comparison:

$$\Delta y_{it} = \theta_0+\alpha_d \Delta CT_{it}+ \boldsymbol{w}_{it}'\boldsymbol{\theta}+\omega_t+\Delta\tau_{it}. \tag{6}$$

Where $\boldsymbol{w}_{it}$ is a set of variables derived from the differences and lags of weather and socioeconomic variables (see Appendix Table A1); θ_0 is the constant term, ω_t is time fixed effect; and $\Delta\tau_{it}$ is the error term. This baseline specification controls for a small set of time-varying county-specific factors, any time-invariant county-specific factors, and national aggregate trends.

I use two sets of control variables for both conventional estimator and double selection estimator to compare the performance of them. For the first set of control variables, I use the variables that were selected based on researchers' knowledge and intuition as a baseline. For the second set of control variables, I cover as many as the climate variables and socioeconomic variables that have been found to be useful to predict yields and CT adoption and rely on the variable selection method to select the variables that are most important.

In the first set of control variables for baseline model (conventional estimator), I include variables that were used in Chen et al. (2021) but replace some weather and socioeconomic variables with similar variables in different form or different regional level. As mentioned in the data section, drought has influence on crop yields, I use the PDSI level in August as the proxy of drought instead of drought indicator created based on PDSI in August used in Chen et al. (2021) to show the year-on-year differences in PDSI. I use the lags of county level fertilizer expenses and petroleum product expenses instead of the lags of national level fertilizer prices and fuel prices because national level data tend to be dropped with time fixed effect. I include the differences of some variables used in Chen et al. (2021) such as GM adoption rate, corn/soybean insured rate, continuous corn/soybean rate, spring wetness and lags of some variables such as fall wetness, price received, CT adoption rate for corn, and CT adoption rate for soybeans (Chen et al., 2021). The discussion of these variables can be found in the previous data section and in the paper of Chen et al. (2021). I use all control variables[2] (before differenced and lagged) mentioned above to perform transformation and interaction to generate technical regressors for post-double-selection estimator.

[2] The lags of CT adoption rates for corn and soybeans are not included in both sets of control variables for post-double-selection estimator.

I add the differences of monthly primary climate variables from April to October, monthly derived climate variables from April to October, GDD between 8 and 30 and above 30 during growing season from April to October, and other production expenses; and the lags of some subcategory incomes and inventory changes to the first set of control variables as the second set of control variables for the baseline model (conventional estimator), I don't include climate variables that are out of growing season because some derived climate variables may exhibit series correlations (Chen et al., 2021), such as PDSI for a given month can reflect the weather condition in previous months (Alley, 1984; Palmer, 1965). Similar as the first set of variables, I use all variables (before differenced and lagged) for conventional estimator to perform transformation and interaction to generate the second set of technical regressors for post-double-selection estimator.

The summary statistics for the dependent and independent variables used in the regressions are reported in Appendix Table 3.

3.2 Robustness check

3.2.1 Overfitting

Overfitting bias is a concern for highly complex fitting methods such as random forests, boosting, and hybrid machine learning methods (Bach et al., 2021), the Double Machine Learning (DML) method (Chernozhukov et al., 2018) developed the cross-fitting concept to overcome overfitting without loss of efficiency. LASSO is less concerned on overfitting, but the DML method can also be used for LASSO estimation. Thus, I adopt the DML method (Chernozhukov et al., 2018) to perform robustness check for the post-double-selection method. The detailed explanation of DML method can be found in Appendix B.

3.2.2 Irrigation counties

For the counties that use irrigation for corn and soybean production, the effects of CT on corn and soybean yields might be different compared with the counties don't use irrigation or use less irrigation. I obtain the total harvested cropland area and the harvested cropland area that is irrigated in the county reported in the 2002, 2007, 2012, and 2017 censuses, and then calculate the proportion of harvested cropland in the county that is irrigated in each census year by dividing the irrigated harvested cropland area by the total harvested cropland area. I identify and drop the counties (132 counties) that have equal to or more than 5 percent of the harvested cropland irrigated in any of the Census year (Kuwayama et al., 2019) to test the effects of CT on corn and soybean yields as robustness checks.

3.2.3 Counties with high residue cover data missing rates

Robustness check is conducted to verify the model performance by excluding states with high missing rates of residue cover data. The states excluded are Iowa, Kansas, Minnesota, and Oklahoma.

4 Results

4.1 Primary Results

I apply LASSO (use R package 'cv.glmnet' (Friedman et al., 2022)) to equations (3) and (2), respectively, to select the lambda to obtain minimum mean squared error (MSE) and select the technical regressors that are important to predict yields and CT adoption rates. The number of folds used for cross validation when implementing LASSO variable selection process is 10; the time fixed effect dummies are excluded from the variable selection, which means that they are included in any cases. After the regressors are selected for predicting yields and CT adoption rates, the union of the technical regressors and the time fixed effect dummies are used as controls to estimate α in the ordinary least squares regression of Δy_{it} on ΔCT_{it}.

The variables selected may vary each time because the samples are randomly split for training data set and test data set when performing cross validation. To make the results more robust, I repeat the variable selection process and regression 1000 times, and follow the median method used in Chernozhukov et al. (2018) to obtain the median estimate and the standard error adjusted for variation across 1000 regressions. The confidence interval is calculated based on the estimate and standard error calculated using median method; adjusted R-squared, RMSE, and control variables excluding time-fixed effect dummies are reported as the ones for the median estimate. The lambda selection process for the median lambda among the 1000 repeats can be found in Appendix Figure A1 and Figure A2.

Table 1 reports the regression results of difference of corn yields on the difference of CT adoption rates applying conventional estimator and post-double-selection estimator to the two sets of control variables in Appendix Table A1, respectively. Regardless of the model specifications and variables used, the results suggest that CT adoption has no significant impact on corn yields. Double selection estimator uses the control variables that captures a cubic trend for the level of crop yield and CT adoption rate, the results of the double selection estimators show a smaller confidence interval and a smaller RMSE compared to the conventional estimators when using the same variables (before transformation and interaction). The less accurate estimation obtained by conventional estimators might be because of potentially incorrect linear specification. If I assume the linear specification is correct, the performance of conventional estimator is improved after adding more variables, this can be seen from the smaller confidence interval and RMSE of conventional (2) compared to those of the conventional (1), thus, the less accurate estimation obtained by conventional estimator (1) can be partially explained by omitted variable biases. The estimate of Double Selection estimator (1) is less accurate compared with that of Conventional estimator (2) even if there are more regressors used for the regression, this might be explained by larger impact of omitted variable bias on the model performance compared with the impact of misspecification when the information included in the model is limited. The number of technical regressors that are selected by double-selection-LASSO for the first set of variables is 267 out of 270 (total number of technical regressors before selection), which means that almost all the regressors provided for selection are important to predict yields or CT adoption. For the second set of variables, 1,539 regressors are selected from 14,262 total regressors. Most technical regressors selected though double selection method are differences and lags of some climate variables and socioeconomic variables with or without interactions with linear or nonlinear time trend, initial differences of some climate variables and socioeconomic variables interacted with linear or nonlinear time trend, climate variables interacted with climate variables and with or without interacted with linear or nonlinear time trend, socioeconomic variables interacted with climate variables or socioeconomic variables and with or without interacted with linear or nonlinear time trend. These selected variables suggest the presence of a nonlinear trend that depends on county-specific characteristics.

Table 2 reports the regression results of difference of soybean yields on the difference of CT adoption rates applying conventional estimator and double selection estimator to the two sets of control variables in Appendix Table A1, respectively. Regardless of the model specifications and variables used, the results suggest that CT adoption has significant and negative impact on soybean yields. The performance of the estimators for soybeans follows the similar pattern as that for corn, double selection estimator using the second set of control variables yields the most accurate estimation. The number of regressors selected for soybeans are similar as that for corn, 264 regressors are selected from 270 regressors for the first set of variables; 1,473 regressors are selected from 14,262 regressors for the second set of variables. There are less regressors selected for soybeans, this is mainly because that less regressors for predicting CT adoption for soybeans are selected. I use the same set of variables for both corn and soybeans, however, there might be some factors important for CT adoption for corn but not for soybeans, or some factors that are important to influence CT adoption for soybeans are not included in the analysis. Similar as corn, different forms and interactions of climate and socioeconomic variables are selected, the selected variables suggest the presence of a nonlinear trend that depends on county-specific characteristics. The coefficient on CT adoption from double selection (2) suggests that a 1% increase in CT adoption can lead to about 0.058% (average effect, 95% Confidence Interval, CI = [0.083%, 0.033%]) decrease in soybean yield, this translate to an average loss of 0.028 bushel/acre (CI = [0.039, 0.016]) given that the average soybean yield across counties and 14 years is 47.47 bu/acre (Appendix Table A2). The average CT adoption rate across counties and years is 55.14%, this means that soybean yields would decline by 3.198% (CI = [4.577%, 1.820%]), this is equivalent to an average loss of 1.518 bushel per acre (CI = [2.173, 0.864]).

4.2 Robustness check

4.2.1 Overfitting check

For the robustness check, I use R package DoubleMLPLR (Bach et al., 2021) to estimate the effect of CT adoption on corn and soybean yields. The number of folds used for cross validation when implementing LASSO is 10; the time fixed effect dummies are excluded from the variable selection, which means that they are included in any cases. I choose a 5-fold cross-fitting and repeat the cross-fitting 200 times, the median of 200 coefficients is reported as the estimate. Row

1 of Appendix Table A4 and A5 show the regression results of using DML method for corn and soybeans, respectively, similar results are obtained. This suggests that overfitting is not a big concern when using double selection method.

4.2.2 Dropping counties with Irrigation effect

Row 2 of Appendix Table A4 and A5 show the regression results of using data dropping counties that have 5% harvested cropland irrigated for corn and soybeans, respectively. Similar effects for both corn and soybeans are obtained, this suggests that the primary results are robust when including counties with more than 5% harvested cropland irrigated.

4.2.3 Dropping states with high residue cover data missing rates

Row 3 of Appendix Table A4 and A5 show the regression results of using data dropping states that have residue cover missing rate for corn and soybeans, respectively. Similar effects for both corn and soybeans are obtained, this suggests that the primary results are robust when including observations with high residue cover data missing rates.

Table 1. Regressions of difference of corn yield on difference of CT adoption.

Estimator	Effect	Robust Std. Error	Confidence Interval 2.5%	Confidence Interval 97.5%	Adjusted R-squared	RMSE	Obs. N	Control Var. N
Conventional (1)	0.027	0.028	-0.028	0.082	0.413	0.201	6501	13
Conventional (2)	0.008	0.018	-0.027	0.043	0.757	0.128	6501	116
Double Selection (1)	0.033	0.025	-0.015	0.081	0.562	0.170	6501	267
Double Selection (2)	-0.000	0.016	-0.031	0.030	0.883	0.078	6501	1539

Notes: The table shows the results from two different estimators using two different sets of variables. Conventional is the model specified with first differencing the variables and using lags of the variables; Double selection is the estimator to apply the post-double-selection method to the partially linear regression model. (1) means using the first set of control variables and (2) means using the second set of control variables listed in Appendix Table A1; for the conventional estimator using the first or second set of control variables, the variables are differenced and lagged control variables, while, for double selection estimator, the variables enter the model after transformations and interactions. Results are obtained by applying the median method (Belloni et al., 2014a) to 1000 repeats with point estimates. The robustness standard errors are calculated by applying the median method to clustered standard errors from the 1000 repeats to adjust for the variation across the repeats. Confidence intervals are calculated based on the reported median estimates and the adjusted clustered standard errors. Other statistics are reported using the statistics for median estimates. Obs. N is the total number of observations used in the regression. Control Var. N is the number of technical regressors (not including time fixed effect dummies) used as controls.

Significant codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

Table 2. Regressions of difference of soybean yield on difference of CT adoption.

Estimator	Effect	Robust Std. Error	Confidence Interval 2.5%	Confidence Interval 97.5%	Adjusted R-squared	RMSE	Obs. N	Control Var. N
Conventional (1)	-0.116***	0.019	-0.154	-0.078	0.262	0.155	6501	13
Conventional (2)	-0.107***	0.016	-0.138	-0.076	0.624	0.110	6501	116
Double Selection (1)	-0.076***	0.018	-0.112	-0.041	0.434	0.133	6501	264
Double Selection (2)	-0.058***	0.013	-0.083	-0.033	0.806	0.070	6501	1473

Notes: The table shows the results from two different estimators using two different sets of variables. Conventional is the model specified with first differencing the variables and using lag of the variables; Double selection is the estimator to apply the post-double-selection method to the partially linear regression model. (1) means using the first set of variables and (2) means using the second set of variables listed in Appendix Table A1; for the conventional estimator using the first and second variables, the variables are used as original, while, for double selection estimator, the variables enter the model after transformation and interaction. Double selection estimator results are obtained by applying the median method (Belloni et al., 2014a) to 1000 repeats with point estimates. The robustness standard errors are calculated by applying the median method to clustered standard errors from the 1000 repeats to adjust for the variation across the repeats. Confidence intervals are calculated based on the reported median estimates and the adjusted clustered standard errors. Other statistics are reported using the statistics for median estimates. Obs. N is the total number of observations used in the regression. Control Var. N is the number of technical regressors (not including time fixed effect dummies) used as controls.

Significant codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

5 Discussions and Conclusions

I find that CT adoption has no impact on corn yield and has significant but small negative impact on soybean yield. The analyses suggest that soybean yields would decline by approximately 3.2% in counties with an average CT adoption rate of about 55%. The finding for corn is consistent with the studies conducted by Chen et al. (2021), and the finding for soybean is slightly different from theirs but is consistent with their results when county-year fixed effect specification is used. The finding for soybeans is also slightly different from that of Deines et al. (2019), the differences might be caused by avoiding some omitted variable bias through adding socioeconomic variables in this study.

Ten percent increase in CT adoption can lead to an average yield loss of about 0.28 bushel/acre (CI = [0.39, 0.16]) for soybeans based on the mean of historical soybean yields across counties from 2005 to 2018. The average harvested acre for soybeans is 82,830 acres (Appendix Table A2) across counties and years, this gives us an average yield loss of about 23,192 bushels (CI = [32,340, 13,253]) for each county. Then the dollar amount loss for each county due to 10% increase in CT adoption for soybeans would be 346,720.40 US dollars (CI = [483,483.00, 198.132.35]) given that the average price received for soybeans for the first five months of Year 2022 is 14.95 $/bushel (USDA-NASS, 2022).

Some farmers adopt CT may be because CT can lower labor and fuel costs, and reduce expenditures on weed management and insecticide (Claassen et al., 2018; Murphy et al., 2006; Uri, 2000), meanwhile, farmers adopting CT can receive government payments though joining the Conservation Stewardship Program (CSP), the Environmental Quality Incentives Program (EQIP), and other cost-share programs. These factors could influence the profits with adopting CT adoption. Farmers may adopt CT purely because of the environmental benefits of CT, the differences of perception of climate change or awareness of soil conservation caused by differences of individual characteristics are hard to be captured at a county level. Other time-varying variables that are hard to be reflected in a county-level data but are important to yield and CT adoption might lead to omitted-variable bias.

The post-double-selection method used in this study provides more accurate estimation than the conventional approach. The attempt to avoid some omitted variable biases using post-double-selection method seems effective in this context. The performance of the post-double-selection is not so well when the information used to explain the dependent and independent variables are limited.

Acknowledgements

This research was conducted as my second-year research paper. I am grateful for comments from my advisor Dr. Ben Gramig, my second-year research paper class professors Dr. Teresa Serra Devesa and Dr. Carol Nelson, and my peers in the second-year research paper class.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 170191. https://doi.org/10.1038/sdata.2017.191

Al-Kaisi, M. M., Archontoulis, S. V., Kwaw-Mensah, D., & Miguez, F. (2015). Tillage and Crop Rotation Effects on Corn Agronomic Response and Economic Return at Seven Iowa Locations. *Agronomy Journal*, *107*(4), 1411–1424. https://doi.org/10.2134/agronj14.0470

Alley, W. M. (1984). The Palmer Drought Severity Index: Limitations and Assumptions. *Journal of Applied Meteorology and Climatology*, *23*(7), 1100–1109. https://doi.org/10.1175/1520-0450(1984)023<1100:TPDSIL>2.0.CO;2

Anandhi, A. (2016). Growing degree days – Ecosystem indicator for changing diurnal temperatures and their impact on corn growth stages in Kansas. *Ecological Indicators*, *61*, 149–158. https://doi.org/10.1016/j.ecolind.2015.08.023

Bach, P., Chernozhukov, V., Kurz, M. S., & Spindler, M. (2021). DoubleML -- An Object-Oriented Implementation of Double Machine Learning in R. *ArXiv:2103.09603 [Cs, Econ, Stat]*. http://arxiv.org/abs/2103.09603

Belloni, A., Chernozhukov, V., & Hansen, C. (2014a). Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies*, *81*(2), 608–650. https://doi.org/10.1093/restud/rdt044

Belloni, A., Chernozhukov, V., & Hansen, C. (2014b). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, *28*(2), 29–50. https://doi.org/10.1257/jep.28.2.29

Bruner, E., & Brokish, J. (2021). *Ecosystem Market Information: Opportunity and Program Comparison [Fact sheet]*. Illinois Sustainable Ag Partnership. https://ilsustainableag.org/wp-content/uploads/2021/09/Ecomarkets-Program-Comparison.pdf

Cabas, J., Weersink, A., & Olale, E. (2010). Crop yield response to economic, site and climatic variables. *Climatic Change*, *101*(3–4), 599–616. https://doi.org/10.1007/s10584-009-9754-4

Chen, B., Gramig, B. M., & Yun, S. D. (2021). Conservation tillage mitigates drought-induced soybean yield losses in the US Corn Belt. *Q Open*, *1*(1), qoab007. https://doi.org/10.1093/qopen/qoab007

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Claassen, R., Duquette, E. N., & Smith, D. J. (2018). Additionality in U.S. Agricultural Conservation Programs. *Land Economics*, *94*(1), 19–35.

Climatology Lab. *TerraClimate*. Climatology Lab. Retrieved July 25, 2022, from https://www.climatologylab.org/terraclimate.html

CTIC. (1992). *National survey of conservation tillage practices*. Conservation Technology Information Center.

DeFelice, M. S., Carter, P. R., & Mitchell, S. B. (2006). Influence of Tillage on Corn and Soybean Yield in the United States and Canada. *Crop Management*, *5*(1), CM-2006-0626-01-RS. https://doi.org/10.1094/CM-2006-0626-01-RS

Deines, J. M., Wang, S., & Lobell, D. B. (2019). Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt. *Environmental Research Letters*, *14*(12), 124038. https://doi.org/10.1088/1748-9326/ab503b

Dixon, B. L. (1994). Estimating Corn Yield Response Models to Predict Impacts of Climate Change. *Journal of Agricultural and Resource Economics*, *19*(1), 58–68.

Dixon, B. L., & Segerson, K. (1999). Impacts of Increased Climate Variability on the Profitability of Midwest Agriculture. *Journal of Agricultural and Applied Economics*, *31*(3), 537–549.

Egli, D. B. (2008). Comparison of Corn and Soybean Yields in the United States: Historical Trends and Future Prospects. *Agronomy Journal*, *100*(S3), S-79-S-88. https://doi.org/10.2134/agronj2006.0286c

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., Qian, J., & Yang, J. (2022). *Lasso and Elastic-Net Regularized Generalized Linear Models*. https://glmnet.stanford.edu, https://dx.doi.org/10.18637/jss.v033.i01, https://dx.doi.org/10.18637/jss.v039.i05

Karlen, D. L., Kovar, J. L., Cambardella, C. A., & Colvin, T. S. (2013). Thirty-year tillage effects on crop yield and soil fertility indicators. *Soil and Tillage Research*, *130*, 24–41. https://doi.org/10.1016/j.still.2013.02.003

Kaufmann, R. K., & Snell, S. E. (1997). A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants. *American Journal of Agricultural Economics*, *79*(1), 178–190. https://doi.org/10.2307/1243952

Kerr, Y. H. (2007). Soil moisture from space: Where are we? *Hydrogeology Journal*, *15*(1), 117–120. https://doi.org/10.1007/s10040-006-0095-3

Kragt, M. E., Dumbrell, N. P., & Blackmore, L. (2017). Motivations and barriers for Western Australian broad-acre farmers to adopt carbon farming. *Environmental Science & Policy*, *73*, 115–123. https://doi.org/10.1016/j.envsci.2017.04.009

Kuwayama, Y., Thompson, A., Bernknopf, R., Zaitchik, B., & Vail, P. (2019). Estimating the Impact of Drought on Agriculture Using the U.S. Drought Monitor. *American Journal of Agricultural Economics*, *101*(1), 193–210. https://doi.org/10.1093/ajae/aay037

Laukkanen, M., & Nauges, C. (2011). Environmental and Production Cost Impacts of No-till in Finland: Estimates from Observed Behavior. *Land Economics*, *87*(3), 508–527.

Lobell, D. B., Cassman, K. G., & Field, C. B. (2009). Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annual Review of Environment and Resources*, *34*(1), 179–204. https://doi.org/10.1146/annurev.environ.041008.093740

Lobell, D. B., Roberts, M. J., Schlenker, W., Braun, N., Little, B. B., Rejesus, R. M., & Hammer, G. L. (2014). Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest. *Science*, *344*(6183), 516–519. https://doi.org/10.1126/science.1251423

Lusk, J. L., Tack, J., & Hendricks, N. P. (2017). *Heterogeneous Yield Impacts from Adoption of Genetically Engineered Corn and the Importance of Controlling for Weather* (Working Paper No. 23519; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w23519

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, Ö., Yu, R., & Zhou, B. (Eds.). (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Mika, J., Horváth, Sz., Makra, L., & Dunkel, Z. (2005). The Palmer Drought Severity Index (PDSI) as an indicator of soil moisture. *Physics and Chemistry of the Earth, Parts A/B/C*, *30*(1–3), 223–230. https://doi.org/10.1016/j.pce.2004.08.036

Murphy, S. D., Clements, D. R., Belaoussoff, S., Kevan, P. G., & Swanton, C. J. (2006). Promotion of Weed Species Diversity and Reduction of Weed Seedbanks with Conservation Tillage and Crop Rotation. *Weed Science*, *54*(1), 69–77.

OpTIS. (2019). *Mapping conservation practices and outcomes in the corn belt* [Final report. a collaborative project between Regrow (formerly Applied Geosolutions LLC and Dagan, Inc.)]. The Nature Conservancy, and the Conservation Technology Information Center.

Palmer, W. C. (1965). *Meteorological Drought* (Research Paper No. 45). Office of Climatology, U.S. Weather Bureau. https://www.droughtmanagement.info/literature/USWB_Meteorological_Drought_1965.pdf

Plastina, A., & Wongpiyabovorn, O. (2021). How to Grow and Sell Carbon Credits in US Agriculture. *Ag Decision Maker Extension.Iastate.Edu/Agdm*. https://www.extension.iastate.edu/agdm/crops/pdf/a1-76.pdf

PRISM Climate Group. (n.d.). Oregon State University. Retrieved July 25, 2022, from https://prism.oregonstate.edu/

Prokopy, L. S., Floress, K., Arbuckle, J. G., Church, S. P., Eanes, F. R., Gao, Y., Gramig, B. M., Ranjan, P., & Singh, A. S. (2019). Adoption of agricultural conservation practices in the United States: Evidence from 35 years of quantitative literature. *Journal of Soil and Water Conservation*, *74*(5), 520–534. https://doi.org/10.2489/jswc.74.5.520

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, *56*(4), 931–954. https://doi.org/10.2307/1912705

Tannura, M. A., Irwin, S. H., & Good, D. L. (2008). Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1147803

Thompson, N. M., Hughes, M. N., Nuworsu, E. K. M., Reeling, C. J., Mintert, J. R., Langemeier, M. R., DeLay, N. D., & Foster, K. A. (2021). *Opportunities and Challenges Associated with "Carbon Farming" for U.S. Row-Crop Producers*. 12.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Toliver, D. K., Larson, J. A., Roberts, R. K., English, B. C., De La Torre Ugarte, D. G., & West, T. O. (2012). Effects of No-Till on Yields as Influenced by Crop and Environmental Factors. *Agronomy Journal*, *104*(2), 530–541. https://doi.org/10.2134/agronj2011.0291

Urban, D. W., Roberts, M. J., Schlenker, W., & Lobell, D. B. (2015). The effects of extremely wet planting conditions on maize and soybean yields. *Climatic Change*, *130*(2), 247–260. https://doi.org/10.1007/s10584-015-1362-x

Uri, N. D. (2000). An evaluation of the economic benefits and costs of conservation tillage. *Environmental Geology*, *39*(3–4), 238–248. https://doi.org/10.1007/s002540050004

U.S. Bureau of Economic Analysis (BEA). (2021). https://www.bea.gov/

USDA-NASS. (2020a). *Cropland Data Layer*. National Agricultural Statistics Service, United States Department of Agriculture. https://nassgeodata.gmu.edu/CropScape/

USDA-NASS. (2020b). *QuickStats*. National Agricultural Statistics Service, United States Department of Agriculture. https://quickstats.nass.usda.gov/

USDA-NASS. (2022). *QuickStats*. National Agricultural Statistics Service, United States Department of Agriculture. https://quickstats.nass.usda.gov/

USDA-RMA. Risk Management Agency, United States Department of Agriculture. https://www.rma.usda.gov/

Van Roekel, R. J., Purcell, L. C., & Salmerón, M. (2015). Physiological and management factors contributing to soybean potential yield. *Field Crops Research*, *182*, 86–97. https://doi.org/10.1016/j.fcr.2015.05.018

West, T. D., Griffith, D. R., Steinhardt, G. C., Kladivko, E. J., & Parsons, S. D. (1996). Effect of Tillage and Rotation on Agronomic Performance of Corn and Soybean: Twenty-Year Study on Dark Silty Clay Loam Soil. *Journal of Production Agriculture*, *9*(2), 241–248. https://doi.org/10.2134/jpa1996.0241

Wu, J., Adams, R. M., Kling, C. L., & Tanaka, K. (2004). From Microlevel Decisions to Landscape Changes: An Assessment of Agricultural Conservation Policies. *American Journal of Agricultural Economics*, *86*(1), 26–41.

Wu, J., & Babcock, B. A. (1998). The Choice of Tillage, Rotation, and Soil Testing Practices: Economic and Environmental Implications. *American Journal of Agricultural Economics*, *80*(3), 494–511. https://doi.org/10.2307/1244552

Yun, S. D. (2022). *AcdcR: Agro-Climatic Data by County for R. R package version 1.0.0.* https://CRAN.R-project.org/package=acdcR

Yun, S. D., & Gramig, B. M. (2019). Agro-Climatic Data by County: A Spatially and Temporally Consistent U.S. Dataset for Agricultural Yields, Weather and Soils. *Data*, *4*(2), 66. https://doi.org/10.3390/data4020066

Zhang, S., Tao, F., & Zhang, Z. (2017). Spatial and temporal changes in vapor pressure deficit and their impacts on crop yields in China during 1980–2008. *Journal of Meteorological Research*, *31*(4), 800–808. https://doi.org/10.1007/s13351-017-6137-z

Appendix A. Tables and Figures

Table A1. Regression variables and data sources

Variables	Definition	Level	Source
Crop yields	Crop produced per area of land, units = bu/acre	County	USDA – NASS (2020b)
CT adoption rates[3, 4]	Percent acres having 30 percent or greater residue cover	County	OpTIS (2019)
Ppt[2] (April - October)	Monthly total precipitation (rain + melted snow), units = mm	County	PRISM database (prism.oregonstate.edu)
Tmax[2] (April - October)	Daily maximum temperature (Averaged over all days in the month) units = C	County	PRISM database (prism.oregonstate.edu)
Tmin[2] (April - October)	Daily minimum temperature (Averaged over all days in the month), units = C	County	PRISM database (prism.oregonstate.edu)
GDD30[2]	Growing degree days over 30 Celsius degree	County	PRISM database (prism.oregonstate.edu)
GDD8_30[2]	Growing degree days between 8 and 30 Celsius degree		
Fall wetness[1, 2, 3, 4] SPEI in September Spring wetness[1, 2] SPEI in April	Standardized Precipitation-Evapotranspiration Index	County	PRISM database (prism.oregonstate.edu)
AET[2] (April - October)	Actual evapotranspiration, liquid water supply plus the soil water utilized, monthly total, units = mm	County	TerraClimate (www. climatologylab.org)
DEF[2] (April - October)	Climate water deficit, the difference between monthly reference evapotranspiration and actual evapotranspiration, monthly total, units = mm	County	TerraClimate (www. climatologylab.org)
Runoff[2, 4] (April - October)	The excess of liquid water supply used by monthly reference evapotranspiration and soil moisture recharge, monthly total, units = mm	County	TerraClimate (www. climatologylab.org)
Soil moisture[2] (April - October)	Water stored in the soil, total column at end of month, units = m^3	County	TerraClimate (www. climatologylab.org)
SRAD[2] (April - October)	Downward surface shortwave radiation, units = W/m^2	County	TerraClimate (www. climatologylab.org)

Variables	Definition	Level	Source
VAP[2] (April - October)	Vapor pressure (Fick & Hijmans, 2017), average for month, units = kPa	County	TerraClimate (www. climatologylab.org)
VPD[2] (April - October)	Vapor pressure deficit, average for month, units = kPa	County	TerraClimate (www. climatologylab.org)
PDSI[2] (April - October)	Palmer Drought Severity Index, at end of month	County	TerraClimate (www. climatologylab.org)
Price received[1, 3]	Average price received for crop in the last year by farmers; Proxy for expected commodity prices	State	USDA-NASS (2020b)
Continuous crop[1, 2] (corn/soybeans)	Percent acres in continuous corn/soybeans, constructed from Cropland Data Layer (CDL)	County	USDA – NASS (2020a)
Fraction insured[1, 2]	Percent acres enrolled in federally subsidized crop insurance program	County	USDA – RMA (www.rma.usda.gov)
GM adoption rates[1, 2]	Percent acres used genetically modified seeds	State	USDA – NASS (2020b)
Cash receipts from marketing[2, 4]	Sum of cash receipts from livestock and crops, units = 1000 dollars	County	BEA (www.bea.gov)
Cash receipts from crops[2, 4]	Gross cash income from all crops sold from the farm or ranch during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Other income[2, 4]	The sum of government payments and imputed and miscellaneous income received during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Government payments[2, 4]	Value of all government (State or Federal) agricultural payments received during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Production expenses[2]	Production expenses during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Seed expenses[2]	See purchased during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)

Variables	Definition	Level	Source
Fertilizer and lime expenses[1, 2, 3, 4]	Fertilizer and lime (including ag. chemicals) purchased during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Petroleum product expenses[1, 2, 3, 4]	Petroleum product purchased during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Hired labor expenses[2, 4]	Hired farm labor expenses during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
All other production expenses[2]	All other production expenses during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Value of inventory change[2, 4]	The sum of value of livestock, crops, and materials and supplies inventory change for during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Value of crops inventory change[2, 4]	Value of inventory change for crops during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)
Value of materials and supplies inventory change[2, 4]	Value of inventory change for materials and supplies during the calendar year, units = 1000 dollars	County	BEA (www.bea.gov)

Notes: This set of data provides monthly data for 631 counties across twelve Corn Belt states from 2005 to 2018.
[1] *means this variable is included in the first set of variables for double selection methods.*
[2] *means this variable is included in the second set of variables for the double selection methods.*
[3] *means this variable is included as lags in the first set of variables for the conventional methods.*
[4] *means this variable is included as lags in the second set of variables for the conventional methods.*
BEA is Bureau of Economic Analysis, NASS is National Agricultural Statistics Services, OpTIS is Operational Tillage Information System, PRISM is Parameter-elevation Regressions on Independent Slopes Model, RMA is Risk Management Agency, USDA is United States Department of Agriculture.

Table A2. Summary Statistics of yields, CT adoption rates, priced received, harvested acres for corn and soybeans and residue cover data missing rate.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Corn yield (bu/acre)	7561	158.05	32.84	19	140.9	180.4	246.7
CT adoption rate, corn × 100	7561	44.59	15.57	0.01	33.59	54.92	100
Price received, corn	7561	4.21	1.29	1.90	3.46	4.76	6.98
Harvested acres, corn	7561	94,358	61,988	600	46,500	132,600	394,000
Soybean yield (bu/acre)	7561	47.47	9.11	10.9	42.1	53.5	82.3
CT adoption rate, soybeans × 100	7561	55.14	15.81	0.01	44.19	66.67	100
Price received, soybeans	7561	10.36	2.35	5.50	9.22	12.3	14.28
Harvested acres, soybeans	7561	82,830	44,530	2,100	48,800	109,000	311,500
Government payment (1,000 $)	7561	6,562	5,975	0	2,808	8,415	79,888
Residue cover missing rate	7561	18.09	25.02	0	2.52	21.25	100

Table A3. Summary Statistics of dependent and independent variables.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Difference of logged yield, corn	6501	0.013	0.263	-1.991	-0.083	0.104	2.129
Difference of CT adoption rate, corn	6501	0	0.147	-0.598	-0.085	0.082	0.623
Difference of logged yield, soybeans	6501	0.012	0.181	-1.124	-0.08	0.104	1.329
Difference of CT adoption rate, soybeans	6501	0.004	0.139	-0.663	-0.081	0.085	0.683

Table A4. Robustness checks for corn.

Estimator	Effect	Robust Std. Error	Confidence Interval 2.5%	97.5%	Adjusted R-squared	RMSE	Obs. N	Control Var. N
DML-lasso	0.015	0.015[1]	-0.015	0.044	-	-	6501	-
Double Selection (1)	-0.004	0.018	-0.039	0.031	0.893	0.076	5133	1454
Double Selection (2)[2]	-0.018	0.021	-0.059	0.023	0.900	0.076	4488	1384

Notes: Regressions of difference of corn yields on difference of CT adoption rate for robustness checks. DML-lasso estimator is used to conduct the robustness check for overfitting. Double selection estimator (1) uses data dropping counties that have 5% harvested cropland irrigated to perform the robustness check. Double selection estimator (2) uses data dropping states that have high residue cover missing rate. DML estimator results are calculated using median method (Belloni et al., 2014a) based on 200 splits with point estimates. Double selection estimator results are obtained by applying the median method (Belloni et al., 2014a) to 1000 repeats with point estimates. The robustness standard errors are calculated by applying the median method to clustered standard errors from the 1000 repeats to adjust for the variation across the repeats. Confidence intervals are calculated based on the reported median estimates and the adjusted clustered standard errors. Other statistics are reported using the statistics for median estimates. Obs. N is the total number of observations used in the regression. Control Var. N is the number of technical regressors (not including time fixed effect dummies) used as controls.

[1] *the standard error for DML-lasso is not robust standard error and it is smaller than the robust standard error and the confidence interval can be bigger and shift a little bit for robust standard error.*
[2] *the number of folds used for cross-validation for this estimator is 5 given much lesser observations.*

Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table A5. Robustness checks for soybeans.

Estimator	Effect	Robust Std. Error	Confidence Interval		Adjusted R-squared	RMSE	Obs. N	Control Var. N
			2.5%	97.5%				
DML-lasso	-0.070***	0.012[1]	-0.095	-0.047	-	-	6501	-
Double Selection (1)	-0.075***	0.014	-0.103	-0.047	0.815	0.065	5133	1398
Double Selection (2)[2]	-0.061***	0.015	-0.091	-0.031	0.832	0.066	4488	1249

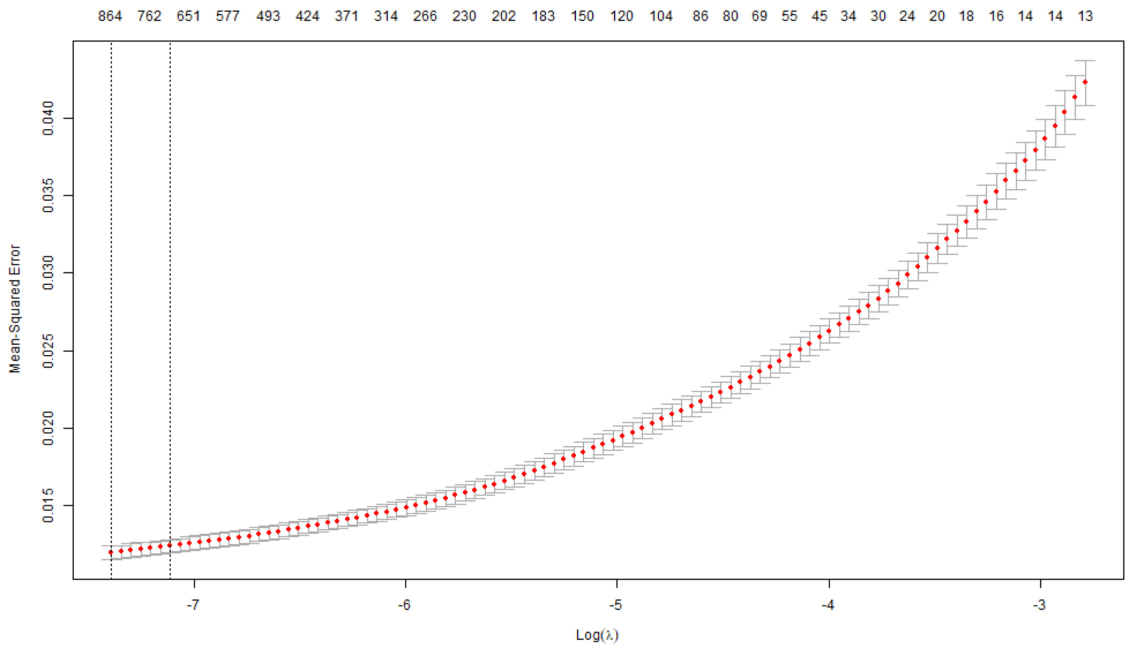
Notes: Regressions of difference of corn yields on difference of CT adoption rate for robustness checks. DML-lasso estimator is used to conduct the robustness check for overfitting. Double selection estimator (1) uses data dropping counties that have 5% harvested cropland irrigated to perform the robustness check. Double selection estimator (2) uses data dropping states that have high residue cover missing rate. DML estimator results are calculated using median method (Belloni et al., 2014a) based on 200 splits with point estimates. Double selection estimator results are obtained by applying the median method (Belloni et al., 2014a) to 1000 repeats with point estimates. The robustness standard errors are calculated by applying the median method to clustered standard errors from the 1000 repeats to adjust for the variation across the repeats. Confidence intervals are calculated based on the reported median estimates and the adjusted clustered standard errors. Other statistics are reported using the statistics for median estimates. Obs. N is the total number of observations used in the regression. Control Var. N is the number of technical regressors (not including time fixed effect dummies) used as controls.

[1] *the standard error for DML-lasso is not robust standard error and it is smaller than the robust standard error and the confidence interval can be bigger and shift a little bit for robust standard error.*
[2] *the number of folds used for cross-validation for this estimator is 5 given much lesser observations.*

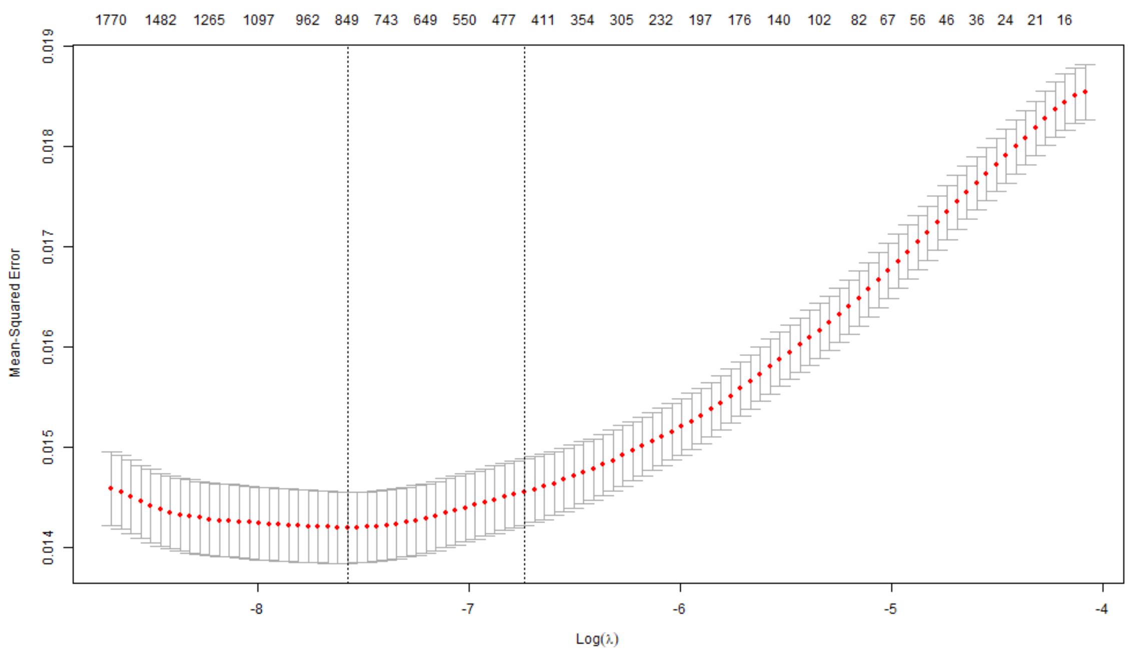
Significant codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure A1. Variable selection curves for corn.

A1a. curve for variable selection for yield



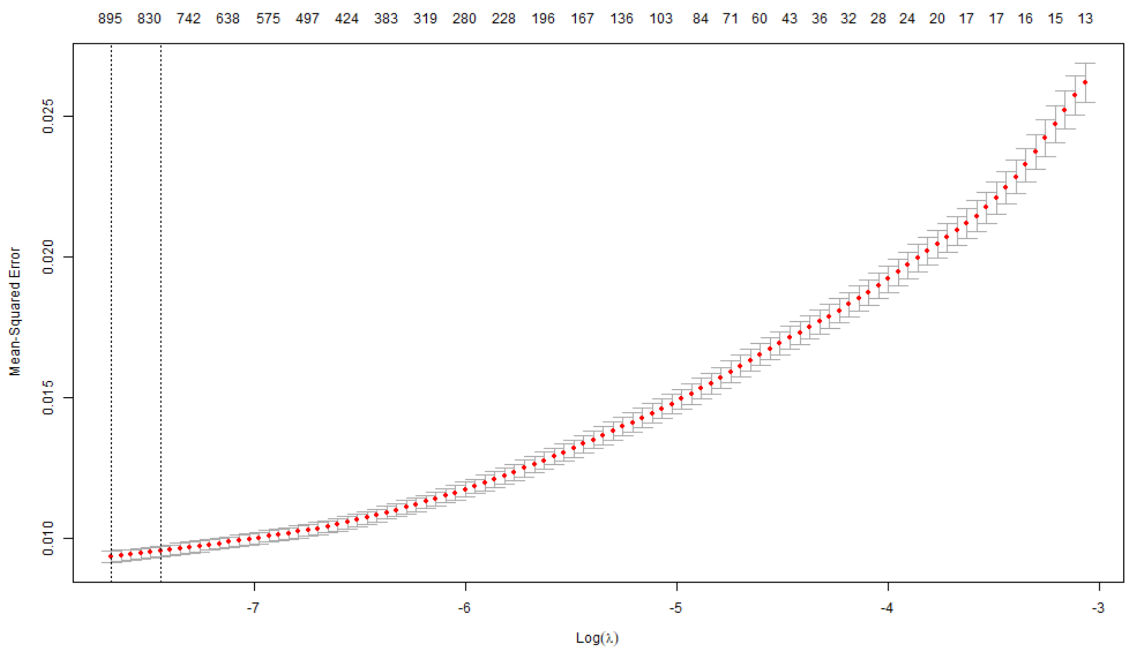
A1b. curve for variable selection for CT adoption



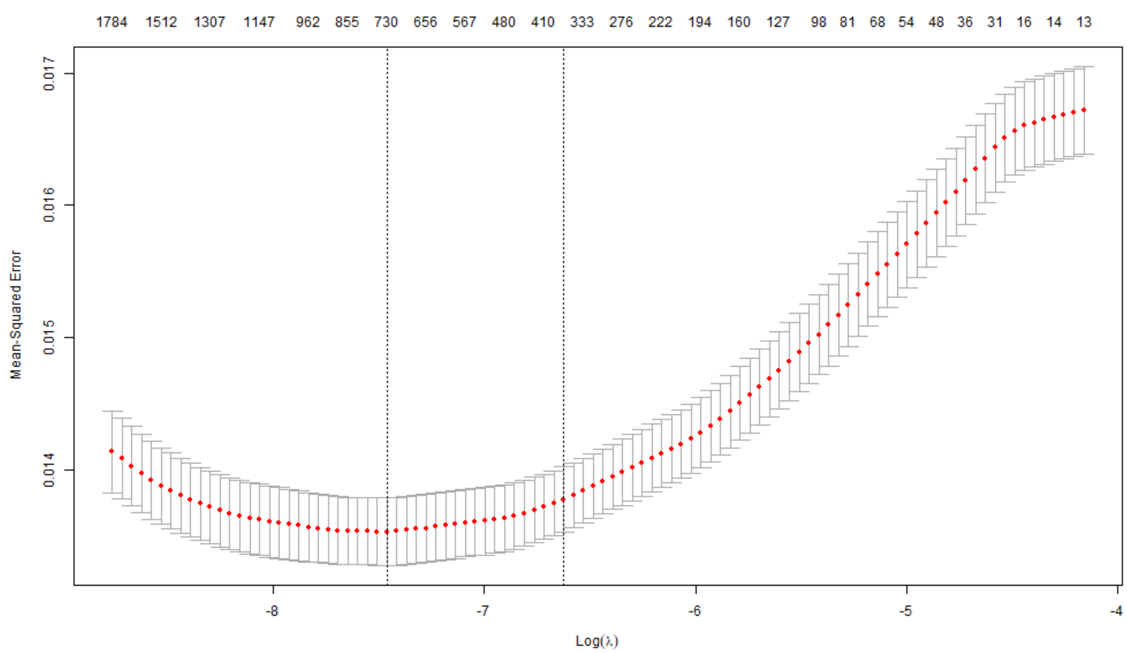
Notes: Figures A1a and A1b show the lambda selection process for the median lambda among the 1000 repeats. The cross-validation curve is shown using red dotted line, the error bars form the upper and lower standard deviation curves along the λ sequence. The left vertical dotted line indicates lambda.min which is the value of λ that gives minimum mean cross-validated error, and the right vertical dotted line indicates lambda.1se which represents the value of λ that gives the most regularized model such that the cross-validated error is within one standard error of the minimum.

Figure A2. Variable selection curves for soybeans.

A2a. curves for variable selection for yield



A2b. curves for variable selection for CT adoption



Notes: Figures A2a and A2b show the lambda selection process for the median lambda among the 1000 repeats. The cross-validation curve is shown using red dotted line, the error bars form the upper and lower standard deviation curves along the λ sequence. The left vertical dotted line indicates lambda.min which is the value of λ that gives minimum mean cross-validated error, and the right vertical dotted line indicates lambda.1se which represents the value of λ that gives the most regularized model such that the cross-validated error is within one standard error of the minimum.

Appendix B. Robustness check details using the DML approach

I apply LASSO to equations (1) and (2) separately to predict Δy_{it} and ΔCT_{it}. This DML approach approximately removes the direct effect of confounding from Δy_{it} through equation (1) and approximately partials out the effect of $\boldsymbol{z}_{it}$ from ΔCT_{it} though equation (2) to obtain the orthogonalized regressor $\Delta\hat{\epsilon}_{it}$, then the estimate of coefficient α can be obtained by using DML estimator (Chernozhukov et al., 2018) as below:

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i\,\in I,t\in T} \Delta\hat{\epsilon}_{it}\,\Delta CT_{it}\right)^{-1}\frac{1}{n}\sum_{i\,\in I,t\in T}\Delta\hat{\epsilon}_{it}\,(\Delta y_{it} - \widehat{\boldsymbol{z}'_{it}\boldsymbol{\beta}} - \mu_t), \tag{B.1}$$

I adopt the *K*-fold cross-fitting approach[3] from Chernozhukov et al. (2018) to split full sample randomly to *K* equal-sized folds, for each $k \in \{1, \dots, K\}$, observations from all other folds (auxiliary sample) are used to apply LASSO and the rest of observations (main sample) are used to estimate the coefficient, the estimator is obtained by averaging the results from *K* estimates. Sample splitting can eliminate the bias induced by overfitting but it can lead efficiency loss, swapping roles of auxiliary and main samples and averaging the results from multiple estimates can regain full efficiency (Chernozhukov et al., 2018).

[3] Sampling splitting doesn't change the prediction structure in the model.