Comparing Deep Neural Network and Econometric Approaches to Predicting the Impact of Climate Change on Agricultural Yield

Michael Keane[†] and Timothy Neal[†]

[†]*University of New South Wales, Sydney, NSW 2000*
E-mail: m.keane@unsw.edu.au
E-mail: timothy.neal@unsw.edu.au

Summary Predicting the impact of climate change on crop yield is difficult, in part because the production function mapping weather to yield is high dimensional and nonlinear. We compare three approaches to predicting yields: (i) deep neural networks (DNNs), (ii) traditional panel-data models, and (iii) a new panel-data model that allows for unit and time fixed-effects in both intercepts and slopes in the agricultural production function - made feasible by a new estimator developed by Keane and Neal (2020) called MO-OLS. Using U.S. county-level corn yield data from 1950-2015, we show that both DNNs and MO-OLS models outperform traditional panel data models for predicting yield, both in-sample and in a Monte Carlo cross-validation exercise. However, the MO-OLS model substantially outperforms both DNNs and traditional panel-data models in forecasting yield in a 2006-15 holdout sample. We compare predictions of all these models for climate change impacts on yields from 2016 to 2100.

Keywords: *Climate Change, Crop Yield, Panel Data, Machine Learning, Neural Net*

1. INTRODUCTION

It is difficult to name a more important public policy issue than the impact of climate change on agricultural yields. But the existing econometric literature that attempts to predict the impact of climate change on yields has produced mixed results. This is not surprising given the inherent difficulty of the exercise. Unfortunately, this means an important component of the cost of future climate change is highly uncertain.

In the conventional econometric approach to modelling yield, weather conditions over a growing season are summarized by a small set of variables. For instance, daily or hourly temperatures are often aggregated into "growing degree days" (GDD) and "killing degree days" (KDD), which summarize the beneficial or harmful temperatures a crop is exposed to over a whole growing season. Standard panel data methods are then used to fit agricultural production functions where yield depends on KDD, GDD and precipitation, along with county and time fixed effects to capture differences in soil quality and technology across counties and time. Finally, climate model predictions of future weather are plugged into the yield model to predict future yield.[1]

One goal of this paper is to assess the potential of deep neural networks (DNNs) to provide better models of crop yield, and hence more reliable predictions of climate change impacts. In contrast to the conventional econometric approach, DNNs can input a huge array of daily weather measures for the entire growing season, along with county and year

---

[1] Papers that pursue this approach include Lobell et al. (2011), Butler and Huybers (2013) and Burke and Emerick (2016). Schlenker and Roberts (2009) use more refined temperature bands.

indicators, and use these to nonparametrically estimate very high-dimensional nonlinear mappings from daily weather to final yields. We compare the performance of DNNs and conventional panel data methods for predicting U.S. county-level corn yield using detailed daily weather data from 1950 to 2015. In a Monte Carlo cross-validation exercise, we find DNNs give more accurate predictions than conventional econometric models. However, we find DNNs perform quite poorly in forecasting for a 2006-15 holdout sample.

The second main goal of the paper is to report results from a new panel data estimator developed by Keane and Neal (2020), called Mean Observation OLS (MO-OLS). This makes it feasible to estimate panel-data models with unit and time fixed-effects in both intercepts and slopes in large panels. This enables us to nest and extend all conventional econometric models in the literature. We find the MO-OLS model achieves a substantial improvement in forecasting performance over standard panel data approaches, both in a Monte Carlo cross-validation exercise and in forecasting for the 2006-15 holdout sample.

A key result is that, in the MO-OLS model, the KDD coefficient is significantly positively correlated with KDD, implying the negative effect of KDD on yield diminishes as KDD increases. We interpret this as due to adaption by farmers to high temperatures.

Finally, we compare predictions of standard panel data models, DNNs and MO-OLS for the impact of climate change on corn yield from the present until 2100. This requires using a climate model to predict future weather conditions in all relevant US counties. We use the Geophysical Fluid Dynamics Laboratory (GFDL) model developed for the US National Oceanic and Atmospheric Administration (NOAA) under the RCP85 CO_2 emissions scenario (which can loosely be described as the 'business as usual' scenario).[2] We incorporate the estimated relationship between the KDD coefficient and KDD (i.e., adaptation) in all forecasts of future yields using the MO-OLS model.

For the conventional and MO-OLS econometric models we forecast future yields both under a "no technical progress" scenario, holding time-fixed effects at their 2015 values, and a "with technical progress" scenario, using VARs to forecast future time-effects. An interesting feature of the DNN is it automatically forms a prediction of technical progress, simply because calendar time is an input variable. We obtain forecasts with and without technical progress by holding the year input fixed at 2015 or letting it evolve.[3]

Historically, U.S. corn yields more than doubled between the 1950s and the present. The conventional panel data models that we consider predict catastrophic drops in yield by 2100 of roughly 60% with technical change and 75% without. The MO-OLS model predicts a stagnation of yields with technical progress, and drop in yield of roughly 40% without. The DNN predictions of future yield trends are similar to MO-OLS.

We view the MO-OLS model predictions as more credible, as the MO-OLS model predicts annual yields in 2006-15 holdout sample far more accurately than either the DNN or conventional panel-data models. Of course, given rapid population growth, even

---

[2]In Keane and Neal (2020) we compare predictions from many climate models and emissions scenarios, to assess the uncertainty across these models/scenarios. That is not our goal in this article. Instead, we want to focus on differences between the econometric and DNN models by comparing their predictions, holding the climate model fixed.

[3]It is in some sense an advantage of the DNN that it implicitly generates its own technical progress forecast, as we don't have to predict technical change ourselves. But it is also a disadvantage, as the nature of the DNN technology prediction is not transparent to an analyst, and we can't control it in order to do scenario evaluations. For example, in an econometrically estimated production function, a positive time trend in the intercept captures neutral technical progress. The DNN offers no such simple interpretation. In this sense the DNN is a "black box."

the stagnation of crop yields predicted by the MO-OLS model (with technical progress) in the U.S. will have serious adverse consequences for world food supply.

In summary, from an econometric perspective, it is impressive that the MO-OLS estimator, which allows for fixed-effects in intercepts and slopes, generates out-of-sample forecasts that are superior to both conventional panel-data models and the far more heavily parameterized DNNs, while maintaining ease of interpretability of the parameters. Clearly it is an important new econometric tool, particularly for large panel datasets.

2. MODELLING AGRICULTURAL YIELD

Crop yield, or agricultural production per acre, is determined annually after the harvest. It is the culmination of many inputs the plants experience over the growing season, including temperature and precipitation, soil quality, technology and farming practices (e.g., seed choice, sowing density, the timing of planting, the use of fertilizers/pesticides, irrigation), and traditional factors of production like capital and labor. We write that:

$$y_{it} = g(\mathbf{T}_{it}, \mathbf{P}_{it}, Q_{it}, A_{it}, \epsilon_{it}) \tag{2.1}$$

where y_{it} is the log of crop yield for county i in year t, $\mathbf{T}_{it} = (T_{i,d=1}, T_{i,d=2}, ..., T_{i,d=D})$ is the history of maximum and minimum temperature on each day d in county i during year t, $\mathbf{P}_{it} = (P_{i,d=1}, P_{i,d=2}, ..., P_{i,d=D})$ is the history of precipitation, Q_{it} is soil quality, A_{it} is technology, and ϵ_{it} captures other inputs. Modelling $g()$ is not straightforward, as it is a high-dimensional nonlinear function. The effect of weather on plant development depends on timing and context, meaning the marginal effect of each element of $g()$ depends not only on its own value but also the values of all the other inputs.

Several recent studies summarize daily temperature data by "degree days" (DDs), i.e., the total time temperature was in certain intervals over a growing season. They then use DDs to predict yield. Most recent studies also include county and year fixed-effects (or time trends) to control for other unmeasured factors that affect yield, such as soil quality, technology, and trends in farming practices.[4,5] For instance, Schlenker and Roberts (2009) estimate regressions of the form:

$$y_{it} = c_i + \eta_s t + \gamma_s t^2 + \sum_{j=0,3,6,..}^{39} \beta_j (DD_{j,it} - DD_{j+3,it}) + \beta_{40} PREC_{it} + \beta_{41} PREC_{it}^2 + \epsilon_{it} \tag{2.2}$$

where $DD_{j,it}$ is the total time the crop experiences temperatures above the $j°$C threshold, s is the State in which county i resides, and $PREC_{it}$ is total precipitation during the growing season. c_i represents county fixed effects, while ϵ_{it} captures idiosyncratic shocks.

Other authors simplify the model by splitting degree days into two intervals, those above and below $29^0 C$, which is a critical threshold for corn.[6] Thus we have beneficial

---

[4]As is well known, input decisions in year t affect soil quality (e.g., nitrogen content, organic content, pathogens) at year $t + 1$, so farmers face a dynamic problem in choosing optimal inputs. This leads to crop rotation (including decisions to leave some land fallow each year). However, the production function in 2.1 is static, because the state variable of current soil quality Q_{it} summarizes the effects of all lagged inputs. Thus, we could ignore dynamics (i.e., effects of lagged inputs) in estimating 2.1 if we observed Q_{it}. Unfortunately we do not observe soil quality at the county/time level, so, as in prior work in this area, we use county and year fixed-effects in an attempt to capture Q_{it}.

[5]Neither we nor the prior literature emphasizes fertilizer as a separate *current* input. It is generally accepted that fertilizer has a cumulative effect on soil quality over time (see Thompson (1963)), so we treat the history of fertilizer use as being captured by Q_{it}.

[6]While moderate temperatures are beneficial, excessively high temperatures can damage corn. For

temperatures (growing degree days) given by $GDD_{it} = DD_{0,it} - DD_{29,it}$ and harmful temperatures (killing degree days) given by $KDD_{it} = DD_{29,it}$. Lobell et al. (2011) and Burke and Emerick (2016) estimate equations similar to the following:[7]

$$y_{it} = c_i + c_t + \beta_1 GDD_{it} + \beta_2 KDD_{it} + \beta_3 PREC_{it} + \beta_4 PREC_{it}^2 + \epsilon_{it} \qquad (2.3)$$

(2.2) and (2.3) can be estimated via one and two-way fixed effects OLS, respectively.

A key shortcoming of (2.2) and (2.3) is they throw away much information about the *timing* and *context* of temperature and precipitation shocks. They do capture one aspect of timing/context by allowing higher temperatures to be a good thing in the cool part of the growing season or in cooler counties, as this increases GDD, while a bad thing in the hot part of the season or in hotter counties, as this increases KDD. But they fail to capture other aspects, for instance that a hot day in June may be less consequential than a hot day at the end of July, even though both may raise KDD by the same amount.[8]

A limitation of (2.3) is that it does not allow for slope heterogeneity across counties. We would expect heterogeneity if farming practices adapt to local climate conditions. E.g., KDD may have a smaller adverse effect on yield in counties with relatively hot climates, if farmers in such counties adopt techniques that make crops less heat sensitive. To accommodate adaptation, Butler and Huybers (2013) use the mean group regression estimator (MG-OLS) due to Pesaran and Smith (1995), as in:

$$y_{it} = c_i + \beta_{1,i} GDD_{it} + \beta_{2,i} KDD_{it} + \beta_{3,i} PREC_{it} + \beta_{4,i} PREC_{it}^2 + \beta_{5,i} t + \epsilon_{it} \qquad (2.4)$$

This can be implemented by running county-specific OLS regressions. Note that this approach includes county-specific time trends, $\beta_{5,i} t$, which is simultaneously more and less restrictive than time fixed effects.[9] A limitation of (2.4), however, is that it ignores the possibility that heat and precipitation sensitivity of the crop may change over time due to changes in technology and soil quality, or due to over-time adaptation of farming techniques (that is common across counties).

To address this limitation, we allow for county and time fixed-effects in *both* intercepts and slopes in an attempt to capture as much variation as possible in the response of yield to temperature and precipitation. Then, we obtain a model of the form:

$$y_{it} = \beta_{0,it} + \beta_{1,it} GDD_{it} + \beta_{2,it} KDD_{it} + \beta_{3,it} PREC_{it} + \beta_{4,it} PREC_{it}^2 + \beta_{5,i} t + \epsilon_{it} \quad (2.5)$$

where:

$$\beta_{kit} = \beta_k + \gamma_{ki} + \lambda_{kt}, k = 0, ..., 4 \qquad (2.6)$$

Leaving the β_{kit} unrestricted gives more parameters than data points, but the structure we place on the fixed-effects in (2.6) enables us to identify the model. The model in (2.5)-(2.6) nests (2.4) if we shut down the time effects λ_{kt} $\forall k$, and also (2.3) as it allows for time effects via the λ_{kt} and the county-specific trends $\beta_{5,i} t$.

One could, in theory, estimate a model with county and time fixed effects in both

---

example, heat increases the rate of transpiration, which drains the plant's water supply, and excessive heat can hamper pollination (see e.g. Lobell et al. (2013) and Tardieu et al. (2018) for details).

[7]Burke and Emerick (2016) compare estimates of (2.3) with those from a "long-difference" specification to assess adaptation. They specify precipitation differently.

[8]Kawasaki and Uchida (2016) allow effects of DDs to differ across three phases of the growing season for rice in Japan.

[9]While equation (2.4) allows for county-specific heterogeneity in time trends, it assumes that all trends are linear. In contrast, (2.3) allows for flexible non-linear time effects (but common to all counties).

intercepts and slopes by applying OLS to the fully interacted specification:

$$y_{it} = \boldsymbol{\beta}' \boldsymbol{x}_{it} + \sum_{j=2}^{N} (\boldsymbol{z}_{it}' d_{ij})' \boldsymbol{\gamma}_j + \sum_{l=2}^{T} (\boldsymbol{x}_{it}' d_{tl})' \boldsymbol{\lambda}_l + u_{it} \tag{2.7}$$

where $\boldsymbol{x_{it}} = (1, GDD_{it}, KDD_{it}, PREC_{it}, PREC_{it}^2)$ and $\boldsymbol{z}_{it} = (\boldsymbol{x_{it}}, t)$ are vectors of covariates, and we also have county dummies $d_{ij} = 1$ if $i = j$ and 0 otherwise, time dummies $d_{tl} = 1$ if $t = l$ and 0 otherwise, and $\boldsymbol{\gamma}_j$ and $\boldsymbol{\lambda}_l$ are vectors of coefficients to be estimated. Then $\hat{\boldsymbol{\beta}}_{it} = \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\gamma}}_i + \hat{\boldsymbol{\lambda}}_t$. Unfortunately, this 'brute force' approach is computationally impractical for large panel datasets such as ours.[10]

However, it is feasible to estimate the model in (2.5)-(2.6) using the "mean observation OLS" (MO-OLS) procedure developed in Keane and Neal (2020). The MO-OLS estimator of β_{it} approximates the OLS estimator to arbitrary accuracy. It is constructed by first running pooled OLS to obtain $\hat{\beta}$, then running regressions by county to collect $\hat{\beta}_i$, then a set of regressions by year to collect $\hat{\beta}_t$. It then constructs the biased preliminary estimator $\hat{\beta}_{it} = \hat{\beta}_i + \hat{\beta}_t - \hat{\beta}$. But Keane and Neal (2020) show how the bias can be calculated to arbitrary accuracy and removed. The expression for the bias is a Cauchy sequence, and as the number of terms grows the OLS estimator is approximated to any desired accuracy.[11]

Note that (2.5) contains $6N + 5T$ parameters and (2.4) contains $6N$, while (2.3) contains only $N + T + 4$. Despite the large number of parameters, estimates from (2.4) and (2.5) are easily interpretable, as the conditional distributions of the slope coefficients can be analyzed to learn about the relationships between weather/climate and yield. Crucially, the MO-OLS estimator allows slope heterogeneity to be correlated with the regressors, so it can flexibly capture nonlinearities in the mapping from weather to yield.

Next, we consider using a deep neural network (DNN) as an alternative to conventional econometric approaches to predicting yield. A DNN is a nonparametric nonlinear model that can predict yield using the vector of all available data $\mathbf{X}_{it} = (i, t, \mathbf{T}_{it}, \mathbf{P}_{it})$, i.e., daily values for max/min temperature, daily precipitation, and year and county dummies. This preserves the information on the timing of weather shocks. Leshno et al. (1993) show that a DNN can approximate any continuous function to arbitrary accuracy. But, while it possesses many attractive features, the DNN is akin to a "black box," where the sheer number of parameters and their nonlinear relation to predicted yield severely limits a researcher's ability to interpret the estimation results.[12]

Figure 1 illustrates the architecture of the DNN. The input layer consists of the set of K inputs $X_{it} = (x_{1,it}, x_{2,it}, ..., x_{K,it})$ specific to county i and year t. The first 'hidden' layer is composed of N_1 neurons. Each neuron takes as input a (different) weighted sum of the outputs from the input layer. Let $M_{1,u} = w_{1,u0} + w_{1,u1}x_{1,it} + ... + w_{1,uK}x_{K,it}$ denote the

---

[10] In our data we have $N = 2209$ counties and $T = 65$ years of data. Thus the "brute force" approach requires $R = (N + T - 1) * (K + 1)$ regressors for NT observations. In our case, this would result in there being $(2209 + 65 - 1) * (4 + 1) = 11,365$ regressors for $\approx 126k$ observations. The size of the regressor matrix renders this "brute force" approach quite impractical. For example, it exceeds the maximum number of regressors allowed in Stata/SE and other software packages.

[11] The MO-OLS estimator can be viewed as an extension of the Frisch-Waugh-Lovell theorem (Frisch and Waugh 1933) to multi-dimensional fixed effects. F-W-L showed the fixed effects estimator can be obtained by demeaning the data for each unit, rather than running OLS using dummies for every unit in the panel. This made fixed effects feasible in practice with very large N. No simple linear transformation of the data yields the MO-OLS estimator, but it can be constructed easily by using the iterative algorithm in Keane and Neal (2020). MO-OLS can also be interpreted as a highly efficient algorithm for solving the linear system $(X'X)\beta = X'Y$ for β while exploiting the structure in (2.5)-(2.6).

[12] Indeed, a DNN is not an "econometric model," as no economic assumptions are involved.

linear index that is input into neuron u in hidden layer 1, where the $w_{1,u0}, w_{1,u1}, ..., w_{1,uK}$ are parameters to be estimated. They are unique to each neuron in each layer.

Figure 1. The Architecture of a Deep Feedforward Neural Network
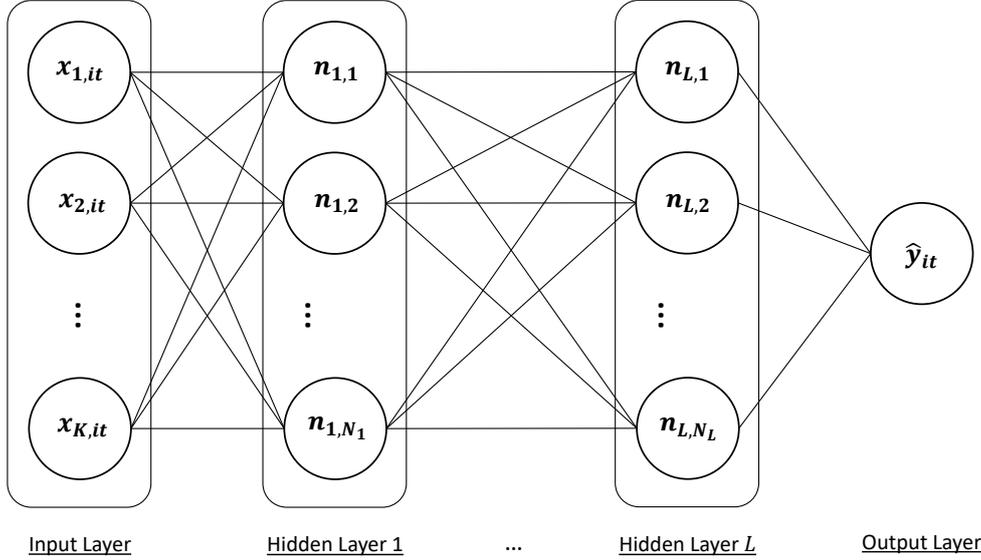
Each neuron u in the first hidden layer plugs the the linear index $M_{1,u}$ it receives into a nonlinear activation function and constructs a scalar output $n_{1,u}$. We use the Exponential Linear Unit (ELU) activation function introduced in Clevert et al. (2015):[13]

$$n_{1,u} = \begin{cases} e^{M_{1,u}} - 1 & \text{if } M_{1,u} < 0 \\ M_{1,u} & \text{if } M_{1,u} \geq 0 \end{cases} \tag{2.8}$$

The vector of outputs from the N_1 neurons in the first hidden layer is $n_{1,1}, n_{1,2}, ..., n_{1,N_1}$.

Neurons in the second hidden layer calculate $M_{2,u} = w_{2,u0} + w_{2,u1}n_{1,1} + ... + w_{2,uN_1}n_{1,N_1}$ and then pass it through the same activation function as in (2.8) to calculate the outputs $n_{2,u}$ for $u = 1, ..., N_2$. The same process is repeated for each hidden layer up to layer L.

Finally, the output layer simply calculates a weighted sum of all the outputs from the neurons of the final hidden layer in order to form the estimate of log yield, $\hat{y}_{it}$:

$$\hat{y}_{it} = w_{L,0} + w_{L,1}n_{L,1} + w_2 n_{L,2} + ... + w_{L,N_L} n_{L,N_L}. \tag{2.9}$$

The number of free parameters (i.e., weights) w_{l,u,N_l} for $u = 1, ..., N_l; l = 1, ...L$ in the DNN model is $(K + 1)N_1 + (N_1 + 1)N_2 + (N_2 + 1)N_3 + ... + (N_{L-1} + 1)N_L + (N_L + 1)$. With such a proliferation of parameters, computing the full set of optimal weights for each neuron poses a significant computational challenge.

LeCun et al. (2012) and Ruder (2016) survey methods for training neural nets, and we adopt many of their suggestions here. We search for weights to minimize the in-sample sum of squared errors $\sum_{t=1}^{T} \sum_{i=1}^{N} (\hat{y}_{it} - y_{it})^2$ using a version of stochastic gradient descent

---

[13]The ELU is a smoothed version of the popular rectified linear unit (RELU) activation function.

called "adaptive moment estimation" (Adam), proposed by Kingma and Ba (2014). It has become very popular in recent DNN applications.

We choose starting values for the weights using the approach of He et al. (2015). Specifically, we draw starting values from a truncated normal distribution with mean 0 and standard deviation $\sigma = \sqrt{2/I}$, where I is the number of inputs to the layer, and any draws more than two standard deviations from the mean are discarded and redrawn. This has the effect of keeping the scale of the input variance to each neuron constant. This helps correct a very common computational problem in DNNs: that the computed gradients for the weights, particularly in the lower layers of the network, can either vanish or explode and significantly slow learning speed.

We use "batch normalization," a popular technique introduced in Ioffe and Szegedy (2015) to to improve training performance. The inputs to neurons inside hidden layers change with the weights of all previous layers, which can lead to computational problems as small changes to weights early in the network may be magnified later in the network. Batch normalization is an operation added prior to the activation function of each neuron. It zero-centers and normalizes each of the neuron's inputs. It then scales and shifts these normalized inputs, where the scaling and shifting parameters are themselves trained as part of the optimization algorithm.

As DNNs have so many parameters, they are very prone to overfit the training data. To reduce this problem, we adopt a technique during training called "dropout." This is due to Hinton et al. (2012) and Srivastava et al. (2014). Each neuron in a specific layer is given a certain probability (here 0.5) that it will not be used at a given iteration. This prevents the weights inside neurons from "cohabitating" with neighboring neurons, which helps prevent over-fitting, thus improving out-of-sample performance. As suggested by Li et al. (2018) we only add dropout to the last hidden layer, as this is considered best practice when applying dropout and batch normalization simultaneously.

As we noted earlier, we use the Adam search algorithm to find the optimal weights. In practice, DNNs never arrive at a single solution, so we stop the algorithm when it fails to find an improvement in a set number of iterations. We implement the model using Google's *Tensorflow* package, which relies on Nvidia's cuDNN library for GPU-acceleration and interacts with the programming language Python. It handles the gradient vectors for each neuron, which are calculated using backpropagation, and then feeds those gradients to the Adam optimization algorithm.

3. MODEL COMPARISON RESULTS

3.1. Data Sources and Methodology

We fit the models of Section 2 using county-level panel data on corn yield and weather. Historical weather data are taken from Schlenker and Roberts (2009). These data contain daily observations on max/min temperature, along with precipitation, for U.S. counties from 1950 to 2015.[14] The daily max/min temperature variables are used to approximate degree day bands $DD_{C,it}$ for $C = 0, ..., 42$ (see Schlenker and Roberts 2009 or Keane and Neal 2020 for details on degree day calculations). These, in turn, are used to construct GDD_{it}, and KDD_{it} for each county i and year t.

We obtain annual corn yield data from the United States Dept. of Agriculture (USDA)

---

[14]Schlenker and Roberts (2009) extended the dataset to 2015 after the publication of their paper, and provided code to map the observations across a grid to each county.

National Agricultural Statistics Service. The data is at the county-level and covers the same 1950-2015 period, although not all counties have crop yield data for all years. Thus we have an (unbalanced) panel with $N = 2,209$ and $T = 65$.[15]

The complete set of models we consider is summarized in Table 1. We classify them as either "Degree day models," which summarize the daily temperature and precipitation data by the GDD, KDD and $PREC$ variables as outlined above, or "Full information models" that include the raw *daily* temperature and precipitation data. The full information models also include the GDD and KDD variables, so as to ensure that they incorporate all available information.[16]

Table 1. Description of Models

Title	Description
	Degree Day Models
County and Time Dummies	FE-OLS without weather data.
FE-OLS	Equation (2.3): FE-OLS with weather data.
Schlenker and Roberts	Equation (2.2): Same as 'FE-OLS' but uses DD intervals of 3 degrees and state-specific quadratic time trends.
Butler and Huybers	Equation (2.4): Mean Group OLS which allows for intercept and slope heterogeneity across counties.
Deep Neural Net	The DNN outlined in Section 2. It has 10 hidden layers, with 3000 neurons in the first layer and 2000 neurons in the others.
MO-OLS	Equation (2.5): Mean Observation OLS as proposed in Keane and Neal (2020). Allows for intercept and slope heterogeneity across counties and over time.
	Full Information Models
LASSO	LASSO with the regularization parameter selected through cross-validation.
Deep Neural Net	The DNN outlined in Section 2. Same structure as above but uses daily weather data as inputs.

We compare the fit of the alternative models using the average mean squared prediction error (AMSE) $\sum_{t=1}^{T} \sum_{i=1}^{N} (\hat{y}_{it} - y_{it})^2$ where y_{it} is log corn yield (in bushels per acre). In Section 3.2 we compare AMSEs in randomly generated holdout samples in a Monte Carlo cross-validation exercise. Then in Section 3.4 we compare AMSEs in a 2006-15 holdout.

---

[15]Following the literature, we exclude counties west of the 100th Meridian that rely heavily on irrigation.

[16]The GDD and KDD variables are complex functions of the daily max/min temperature data. They are good summary statistics because the scientific literature shows that roughly 29 degrees is a critical temperature threshold for corn. In theory the DNN could figure this out and construct GDD and KDD itself. But in practice it might give the econometric models an unfair advantage to incorporate this *a priori* information. By letting the DNN use the GDD and KDD variables, we insure a fair comparison.

3.2. Monte Carlo Cross-Validation Results

Here we compare out-of-sample fit of the competing models in a Monte Carlo cross-validation exercise. We randomly partition the data, with 80% of the (i, t) observations chosen for use in estimation (the training or in-sample data), and the remaining 20% held-out and predicted using the trained model (the testing or out-of-sample data). To avoid accidentally selecting a particularly hard/easy to fit testing sample, we repeat this procedure using multiple random partitions, and report the average results.[17]

Table 2 presents both in-sample and out-of-sample fit results for all models described in Section 2, presented in the same order as in Table 1. The first column shows results for the full sample, while the remaining columns report results for subsamples. The first row presents a simple baseline with only county and time effects. This model gives an R^2 of 0.79 in-sample (0.78 out-of-sample), showing the importance of county/time effects.

The next three rows present results for three traditional econometric models that are well-known in the current literature: FE-OLS based on the GDD/KDD variables, FE-OLS based on $3^oC\ DD$ intervals, and MG-OLS based on the GDD/KDD variables. These correspond to equations (2.3), (2.2) and (2.4), respectively, and to the models of (i) Burke and Emerick (2016) and Lobell et al. (2011), (ii) Schlenker and Roberts (2009), and (iii) Butler and Huybers (2013), respectively.

The FE-OLS and Schlenker-Roberts models, which rely on GDD/KDD and $3^oC\ DD$ intervals, respectively, give nearly identical fits (as measured by AMSE) both in-sample and out-of-sample. Thus, the use of GDD/KDD vs. more refined temperature intervals seems to make little difference. Both models give out-of-sample AMSEs of 0.059 compared to 0.071 for the baseline error components model.

The Butler-Huybers (MG-OLS) model, which allows for county fixed-effects in intercepts and slopes, as well as county-specific time trends, gives a substantial improvement in in-sample AMSE (i.e, 0.040 compared to 0.056 for FE-OLS and 0.068 for county/time effects). However, the out-of-sample AMSE is 0.058, which is scarcely an improvement over FE-OLS. Thus, there is evidence of over-fitting for the MG-OLS model.

The fifth row of Table 2 reports results for a simplified DNN based only on the GDD/KDD and $PREC$ data (i.e., not using daily weather data). The in-sample AMSE is 0.026, which is a substantial improvement over the traditional econometric models. We use 10 hidden layers because this was needed for the DNN based on DD variables to give a better in-sample AMSE than the MO-OLS model (see below).

As expected, the DNN's out-of-sample fit is worse than its in-sample fit. Its out-of-sample ASME is twice as large, at 0.053, giving evidence of over-fitting. Nevertheless, this is a clear improvement over the 0.058 to 0.059 values given by the conventional econometric models. As this version of the DNN is not using additional data, this improvement in fit arises solely because it better captures the nonlinear functional form that maps the $i, t, GDD_{it}, KDD_{it}, PREC_{it}$ variables to annual corn yield.

The sixth panel of Table 2 reports results using MO-OLS. Its in-sample AMSE of 0.030 is 25% lower than the MG-OLS model it nests, suggesting the time fixed effects that MO-OLS adds are very important for fitting the data. These time effects capture changes in technology over time that are not captured by simple linear trends.

---

[17]All sub-samples used in estimation contain some observations for each county i and year t. This enables us to estimate a complete set of county and year fixed effects. As all models include county/year effects, they are being compared on how well they predict idiosyncratic fluctuations in yield at the (i, t) level.

Table 2. Monte Carlo Cross-Validation Results

Estimator	Average Mean Squared Error					
	Total	Low KDD	High KDD	Corn Belt	Northern	Southern
Degree Day Models						
County and Time Dummies						
in sample	0.068	0.039	0.100	0.044	0.071	0.079
out of sample	0.071	0.041	0.105	0.046	0.074	0.083
FE-OLS						
in sample	0.056	0.031	0.085	0.033	0.059	0.068
out of sample	0.059	0.033	0.089	0.035	0.061	0.072
Schlenker and Roberts						
in sample	0.057	0.040	0.076	0.034	0.060	0.067
out of sample	0.059	0.041	0.080	0.035	0.063	0.071
Butler and Huybers						
in sample	0.040	0.025	0.057	0.027	0.044	0.044
out of sample	0.058	0.035	0.083	0.040	0.062	0.065
Deep Neural Net						
in sample	0.026	0.016	0.036	0.016	0.018	0.036
out of sample	0.053	0.029	0.077	0.033	0.043	0.071
MO-OLS						
in sample	0.030	0.018	0.044	0.017	0.031	0.037
out of sample	0.045	0.026	0.066	0.026	0.044	0.055
Full Information Models						
LASSO						
in sample	0.044	0.025	0.063	0.025	0.045	0.054
out of sample	0.047	0.027	0.067	0.026	0.048	0.058
Deep Neural Net						
in sample	0.002	0.002	0.004	0.001	0.002	0.003
out of sample	0.028	0.014	0.041	0.010	0.021	0.037

Note: This table contains Monte Carlo cross-validation results for the models and data outlined in Section 3.1 and Table 1. In each iteration, a random 20% of the data is held out. All results are averages across 100 Monte Carlo iterations, except the Deep Neural Net results which are averages of 10 Monte Carlo iterations.

A more striking finding is that the out-of-sample fit of the MO-OLS model is clearly superior to all competing models. Its AMSE of 0.045 is 15% lower than the DNN, and 25% lower than MG-OLS. Thus, the rich pattern of fixed-effects in both intercepts and slopes allowed by MO-OLS enables it to more accurately capture the nonlinear mapping from weather to crop yield. For instance, as we discussed in Section 2, allowing for county/time fixed-effects in the KDD coefficient enables the model to capture the pattern that the negative effect of high temperatures is less in both hotter counties and time periods. We interpret this as due to adaptation (see Keane and Neal (2020) for discussion).

Next we turn to the bottom of Table 2, which reports results for the full information models that use 3307 covariates.[18] The seventh row of Table 2 reports LASSO results. Interestingly, the LASSO model provides a substantially better fit than the two FE-OLS

[18] 365 values for daily max/min temperature and rainfall, 2209 county dummies, GDD, KDD, and year.

models. It does not fit quite as well in-sample as the Bulter-Huybers (MG-OLS) model, but it fits much better out-of-sample - i.e., there is no evidence of over-fitting for LASSO. The out-of-sample AMSE of the LASSO model (0.047) is slightly worse than the MO-OLS model (0.045), but it beats all other DD-based models. The fact that LASSO does this well suggests the timing of weather variables through the growing season is important.

The last row of Table 2 reports results from the DNN that uses the full complement of daily max/min temperature and rainfall data. It generates a near perfect in-sample fit (AMSE = 0.002). More importantly, the out-of-sample AMSE is only 0.028, which is 45% smaller than the DNN that only uses the annual degree day data, and 38% smaller than the MO-OLS model. Furthermore, it is more than twice as accurate as the conventional econometric models (FE-OLS and MG-OLS). Thus, the DNN seems able to learn a great deal about how the exact timing of temperature and rainfall during the year affects yields, and it exploits this to produce far more accurate predictions than models that rely on aggregate summary statistics.

To summarize our results so far, the ranking of models in terms of out-of-sample AMSE is (1) the DNN using the daily weather data, (2) the MO-OLS degree day model, (3) LASSO using daily data, and (4) the DNN degree day model. These all perform substantially better than the conventional econometric models (FE-OLS and MG-OLS). Note that we do not implement MO-OLS on the daily weather data, as a model with fixed-effects in slopes for so many covariates would exhaust the degrees of freedom in the data. However, it may be possible to estimate MO-OLS models that add a few additional covariates in an attempt to capture some of the information on timing of weather variables. We leave this for future research.

An important way to assess model specification is to consider fit in non-random subsamples. If a model fits poorly in certain subsamples, it may reveal dimensions in which it is misspecified. Thus, the right five columns of Table 2 compare model fit: (i) in cases where KDD was above/below average (i.e., hot vs. cold counties/years), and (ii) in different regions of the country - the Corn Belt, Northern States and Southern States.

Every model has an out-of-sample AMSE for high KDD cases that is much worse than for low KDD cases. Thus, all models have greater difficult in accurately predicting yields in hot counties/years. The greatest divergence is for the FE-OLS model; it has a relatively good AMSE of 0.033 in low KDD cases, but the worst AMSE of all models in high KDD cases (0.089). The DNN model based on daily data gives the lowest AMSE in both low and high KDD cases, while the MO-OLS model is second best in both cases.

Turning to the geographic breakdown, all models find it easiest to predict yield in the Corn Belt, followed by the Northern States and lastly the Southern States. This is presumably because the Corn Belt has the most consistently favorable growing conditions, making it easier to predict yields. The DNN using daily data has the best out-of-sample AMSE in all three regions. The MO-OLS model is tied for second with LASSO in the Corn Belt, in third place slightly behind the DNN based on degree day data in the North, and clear second in the South. In summary, the superior performance of the DNN, MO-OLS and LASSO models over the conventional econometric models is a consistent pattern across both hot/cold counties and time periods, and across geographic regions.

3.3. Interpreting the Neural Network Results

In the previous section we found that the DNN using the daily weather data predicted yields much more accurately than the DNN using only annual degree data data. This

implies that the exact timing of weather shocks over the year is important for predicting yield. Yet, given the extremely large number of parameters in the DNNs, how do we learn about the impact of specifically timed weather shocks on yield?

After estimating a MO-OLS model, it is possible to analyze the distribution of the fixed-effects associated with each covariate to gain insight into the structure of the non-linear relationship between weather and yield. In a DNN, however, one can only compute marginal effects of covariates numerically. These marginal effects depend not only on the county, year, and day of the shock, but also the values of all other covariates in that county/year. Accordingly, in this section we compute marginal effects using perturbed historical temperature paths that increase maximum or minimum temperature by 1 degree in a particular county/year/day, while holding all other covariates fixed.

Figure 2 reports the average marginal effects of *ceteris paribus* changes in daily max/min temperature across all counties and years. The horizontal axis shows the day of the year when the temperature shock occurs, and the vertical axis shows the effect on annual log corn yield. The solid (dashed) line shows the effects of a unit increase in the daily maximum (minimum) temperature. Note that in the U.S. corn is planted in April-May, pollination ('silking') and 'grain filling' occurs in July-August, and the harvest is in September-October. The pace of physical growth of the plant, and the rate of uptake of nutrients, increases rapidly during July, and is greatest in August.

Figure 2. Marginal Effect of Temperature shocks on Yield by day of year
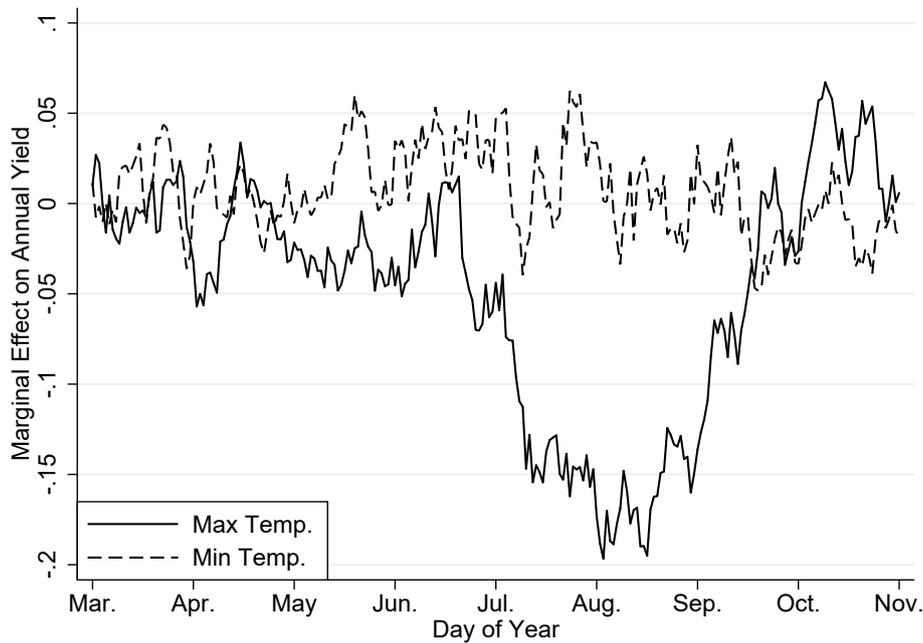
Note: This figure plots the average marginal effects across counties and time of an increase by one degree of maximum or minimum temperature by day of the year.

Consistent with intuition, Figure 2 shows that increases in daily maximum temperature have much larger negative effects on annual yield if they happen during July/August,

which is the hottest part of the growing season and coincides with pollination and grain filling (when the plant grows most rapidly). During that period, a 1 degree Celsius increase in maximum temperature on just one day reduces annual yield by an average of 0.17%.[19] The effects are much smaller earlier and later in the growing season.[20] Thus, the DNN generates results that appear consistent with scientific evidence on the effect of temperature on the corn plant over the life-cycle.[21]

3.4. 2006-15 Holdout Sample Results

Here we compare out-of-sample fit of the competing models using the last ten years of data as a holdout sample. We fit (or train) the models using 1950-2005 data, and then forecast county level yields for 2006-15. To predict holdout sample yields using the econometric models we need to forecast future values of the time varying parameters that capture technical progress.

For the FE-OLS model we must forecast the future values of the year effects c_t in the intercept of equation (2.3). We make these forecasts using AR(1) models fit to the historical year-effect estimates.[22] These models include only the lagged time effect, along with time trends. As the dependent variable is log yield, in specifying the time trends we rule out using t and t^2, as that would permit exponential yield growth in levels. Instead, we capture trends using the two variables $log(t)$ and $(log(t))^{1/2}$.

The MO-OLS model in equations (2.5)-(2.6) contains a vector of five time effects. We forecast these using a VAR(1) system that includes the vector of lagged time effects along with the $log(t)$ and $(log(t))^{1/2}$ variables in each equation. The estimates of all the AR(1) and VAR(1) models are presented in the online Appendix.

The Schlenker-Roberts and Butler-Huybers (MG-OLS) models in (2.2) and (2.4) capture technical progress using time trends, so forecasting time effects is trivial. The same is true of the LASSO model. And the DNN generates its own forecast of future time effects simply by letting the year input evolve over time. This is an advantage of the DNN, but also a weakness, as it is hard to interpret what the implicit time trend captures.

Table 3 reports the fit of the competing models to the 2016-15 holdout sample. The results are strikingly different from those in Table 2. Here, the DNN that uses daily weather data generates the least accurate forecasts of all competing models - even worse than the simple benchmark that includes only county and time effects. The LASSO model generates the next worst forecasts, followed by the DNN using the DD variables. The MO-OLS model generates the most accurate forecasts of all competing models. Its AMSE of 0.065 is 22% lower than the closest competitor, which is the FE-OLS model.

The holdout sample contains the severe drought of 2012, and the right two columns of Table 3 focus on that year. The MO-OLS model again generates the lowest AMSE of all competing models, 30% smaller than the nearest competitor. And if we look at the

---

[19]Note that this is only the average across counties and time, and the actual predicted marginal effect will vary with the values of the other covariates.

[20]Increases in minimum temperature have much smaller effects and lack any clear pattern.

[21]Notably, in the literature on heat effects on yield, one strand emphasizes differential effects by phase of the growth cycle (e.g., classic papers by Wallace (1920) and Thompson (1988) and more recently Tannura et al. (2008)), and notes the importance of July/August temperature, while another uses cumulative degree days over the growing season, and emphasizes non-linearity in the effect of heat (e.g., Schlenker and Roberts (2009)). The DNN can handle both features simultaneously.

[22]We use the same procedure to forecast future time effects in the benchmark county/time effect model.

Table 3. Out-of-sample forecasting performance with 2006-15 Hold Out

Estimator	Mean Squared Error		Mean Bias
	2006-15	2012 Drought	
Degree Day Models			
County and Time Dummies	0.101	0.251	19%
FE-OLS	0.083	0.167	5%
Schlenker and Roberts	0.090	0.151	-4%
Butler and Huybers (MG-OLS)	0.086	0.162	21%
Deep Neural Net	0.103	0.149	-2%
MO-OLS	0.065	0.104	-1%
Full Information Models			
LASSO	0.119	0.172	5%
Deep Neural Net	0.125	0.201	10%

Note: This table contains results of the models and data outlined in Section 3.1 and Table 1. All models use 2006-15 observations as its holdout sample.

prediction of average national yield during the drought year, the MO-OLS model is off by only 1%, making it the most accurate of all models.

Figure 3. Averaged Predicted Corn Yield 2006-15 by Model

Note: This graph presents out-of-sample predictions of average corn yield from 2006 to 2015. The results are presented as a weighted average across U.S. counties.

Figure 3 reports actual and forecast values of annual average corn yield (over all counties) in bushels per acre over the 2006-15 period. Note that the MO-OLS model tracks average yield quite accurately over the whole period. In contrast, the DNN using all weather data generates a national forecast that is far too smooth - it overpredicts yield by 10% in 2012, while underestimating yield substantially in good years (2009, 2013-5). The FE-OLS model's annual yield predictions fall roughly in between.

To avoid clutter in Figure 3, the results for other models, which generally do worse, are reported in the online Appendix. It is worth noting that the DNN based on DD variables misses yield in 2012 by only 2%. However, it understates yield by 25% to 30% in the last three years of the holdout (2013-5), making it appear quite unreliable for forecasting yield further into the future. A smaller DNN with three hidden layers does no better.

In summary, the MO-OLS model clearly fits the 2006-15 holdout data much better than the DNNs, and also better than the conventional panel-data models. It is the only model that performed well in both the MC cross-validation exercise and in the 2006-15 holdout. Based on this we conclude it is the most reliable of the competing models for forecasting future yields.

4. PREDICTING YIELD UNDER CLIMATE CHANGE

4.1. Methodology

Here, we compare predictions of the models of Sections 2 and 3 for annual corn yield from 2016 to 2100. To predict future yields, the models require forecasts of temperature and precipitation in each corn growing county in the U.S. through to 2100. A global climate model (GCM) can provide these predictions if given a CO_2 emissions scenario. We use the NOAA's GFDL-CM3 model.[23] For emissions, we use 'representative concentration pathway' RCP85, which leads to a radiative forcing value (i.e. the balance between incoming and outgoing solar radiation) in 2100 that is 8.5 times pre-industrial levels. This reflects strong emissions growth, and can be loosely described as the "business as usual" scenario. Given this scenario, the GCM provides forecasts of daily max/min temperature and precipitation for each corn growing county from now to 2100. We construct GDD and KDD from these data in the same way as in the historical data.

In Keane and Neal (2020) we analyze the variability of forecasts over several different CO_2 scenarios and many climate models. But our focus here is on comparing the behavior of the different econometric and machine learning models, so we consider only the RCP85 CO_2 scenario and the NOAA's GFDL-CM3 climate model. In particular, given that our MO-OLS model with county/time fixed effects in intercepts and slopes dominates both the DNNs and the traditional panel-data models in our out-of-sample fit tests, it is interesting to see if it generates much different forecasts of future yield.

As we noted in Section 3, the MO-OLS model contains five time varying parameters that must be forecast to predict future yields. However, in Keane and Neal (2020) we find that the coefficients on precipitation exhibit no significant variation over time in the historical data, so we decided to hold them fixed at 2015 values when forecasting future yields. In Keane and Neal (2020) we also find that the KDD coefficient is very well approximated by the nonlinear function $\hat{\beta}_{2,it} = 0.0025(log(KDD_{it})) - 0.0183$, obtained

---

[23]This GCM was developed for the IPCC's Coupled Model Intercomparison Project version 5 (CMIP v5). The GCM predictions were converted from a grid to average county values (i.e. downscaled) using the BCCA procedure (bias corrected constructed analogs) by Reclamation (2013).

by regressing estimates of $\hat{\beta}_{2,it}$ from equation (2.5) on $log(KDD_{it})$ in our sample period. We interpret this relationship as arising from adaptation of farming practices to rising temperature, such that sensitivity of yield to KDD falls as KDD increases.

Keane and Neal (2020) also find significant variation in the intercept and GDD coefficient over time. But, in contrast to the KDD coefficient, they find no systematic pattern that can be used to forecast future intercept and GDD coefficient values. Thus, we forecast $\hat{\lambda}_{0t}$ and $\hat{\lambda}_{1t}$ in equation (2.6) using a VAR estimated on the historical values obtained from estimates of equations (2.5)-(2.6). Specifically, we use a VAR(1) system of two equations, which include first lags of $\hat{\lambda}_{0t}$ and $\hat{\lambda}_{1t}$ along with $log(t)$ and $(log(t))^{1/2}$ to capture time trends. The VAR results are reported in the online Appendix.[24]

To summarize, we forecast yields with technical progress in the MO-OLS model using:

$$\hat{y}_{it,MO} = \hat{\beta}_{0it} + \hat{\beta}_{1it}GDD_{it} + \hat{\beta}_{2it}KDD_{it} + \hat{\beta}_{3i,t=2015}PREC_{it} + \hat{\beta}_{4i,t=2015}PREC_{it}^2 + \hat{\beta}_{5i}t$$

for $t = 2016, ..., 2100$ where $\hat{\beta}_{kit}$ for $k = 0, .., 2$ are forecast as described above. We already discussed in Section 3.3 how we forecast the time effects in the other models.

It is also interesting to predict the effects of climate change alone, holding technology fixed. To do this using the econometric models, we simply hold all time fixed-effects at their 2015 estimated values (or, in the case of the Schlenker-Roberts and Butler-Huybers models, we fix the time trend at 2015). Using the MO-OLS estimates, we can forecast a scenario where adaptation occurs, but other forms of technical progress are shut down. To do this we simply set $\hat{\beta}_{0it} = \hat{\beta}_{0i,t=2015}$ and $\hat{\beta}_{1it} = \hat{\beta}_{1i,t=2015}$ for $t > 2015$, while letting $\hat{\beta}_{2it}$ vary with KDD as discussed above.

As we discussed earlier, the DNN automatically predicts technical progress, simply because time is one of the input variables. The difficulty with the DNN is that, in contrast to the econometric models, it is not clear how to shut down technical change to predict the effects of climate change alone. One idea would be to hold the time input variable fixed at, say, 2015. The problem with this idea is that the DNN may use the time variable to capture temporal changes other than technical progress (e.g., adaptation, changes in unmeasured inputs, changing aspects of climate not captured by other variables, and many other factors that may be unknown to us).[25] This contrasts sharply with the econometric models, where the time variation in intercepts and slopes has a clear technological interpretation.[26]

Nevertheless, for comparison with the econometric models, we also present forecasts from the DNN both with the time variable allowed to update and with the time variable held fixed, but recognizing that the latter is likely to shut down not only technical progress, but also other factors that have affected yield over time.

We present results as the actual yield value, not the logarithm, by simply applying the $e^{\hat{y}_{it}}$ transformation to the projected values. Results are presented as a weighted

---

[24]Recall we forecast the 2006-15 holdout by letting the vector of all five time effects in the MO-OLS model follow a VAR. Here we forecast only two of the time effects using a VAR. We felt it was advisable to incorporate more structure when forecasting all the way to 2100. Hence, given that time effects in the precipitation coefficients appear insignificant, we hold them fixed at 2015 values. And given that the KDD coefficient has a strong relationship with KDD itself, we impose that relationship in forecasting.

[25]Conversely, the DNN may even use trends in the other inputs (like the weather variables) to help capture technical change. The fundamental problem is that the DNN is a 'black box' and while it forecasts very well we do not fully understand the underlying mechanisms at work.

[26]Although, as we noted earlier, we would also expect the time effects to capture changes over time in soil quality, which may in part be induced by lagged land and fertilizer usage.

average across counties, where the counties are weighted by their historical average corn production (in order to better estimate the national average corn yield).

4.2. Future Yield Forecast Results

Figure 4 presents predictions of corn yield from 2016 to 2100 from: (i) the conventional FE-OLS model, (ii) the MO-OLS model, and (iii) the DNN based on daily weather data.[27] These forecasts hold technology fixed at 2015 levels in the econometric models, and hold time fixed at 2015 in the DNN, in an attempt to gauge the impact of climate change alone in the absence of technical progress.

The dashed line in Figure 4 is the average KDD path (across all counties) predicted by NOAA's GFDL-CM3 model under the RCP85 emissions scenario. Note that the model forecast is deterministic - it can be viewed as a single representative path, containing future good years and drought years. The model predicts that KDD will trend upward over time, and also become much more volatile (particularly after 2060).

Figure 4. Prediction of Corn Yield under Climate Change with No Technological Progress
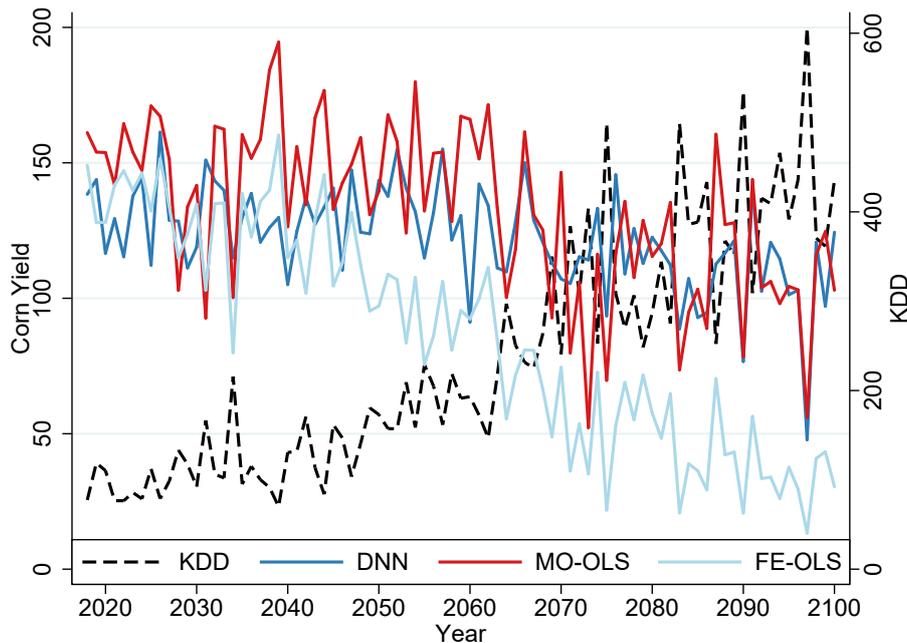
Note: This graph presents predicted average corn yield using the RCP85 emissions scenario and the GFDL climate model. The results are presented as a weighted average across U.S. counties, and time is held fixed at 2015 in each model.

The FE-OLS model predicts catastrophic consequences of climate change. It predicts corn yields will drop by about 70% by 2100, bringing them back to 1940s levels. The MO-OLS and DNN models predict very similar - and somewhat less catastrophic - drops in yields. They both predict drops of about 40% by 2100, which would bring yields back

---

[27] As the three traditional economic models generate very similar results, we report only the FE-OLS model to avoid cluttering the graph.

to levels last seen in the 1970s and 80s. The key difference is that the MO-OLS model predicts that yields will become much more volatile as KDD becomes more volatile, while the DNN predicts much smoother yields over time. Recall that the DNN forecasts for the 2006-15 holdout exhibited "excessive smoothness," while the MO-OLS model predicted annual yield fluctuations quite accurately. This same pattern is clearly reflected here.

Figure 5 reports results that account for technical progress. The FE-OLS model implies technical progress will do very little to mitigate catastrophic drops in yields. Even with technical progress, it predicts yields will fall to about 50 bushels per acre by 2100 (i.e., 1950s levels). The Schlenker-Roberts and MG-OLS results (not shown) are similar.

In contrast, the MO-OLS and DNN models both predict that technical progress will largely counteract the downward trend in yields induced by climate change, and that corn yield will grow (very) slightly over the next two decades and then flatten out. It is worth noting, however, that such a stagnation in the historical rapid growth in US corn yields is likely to create very serious problems for world food supply with a growing population. Thus, even a 'stagnation' scenario is, in effect, disastrous.

Figure 5. Projection of Corn Yield under Climate Change with Technological Progress
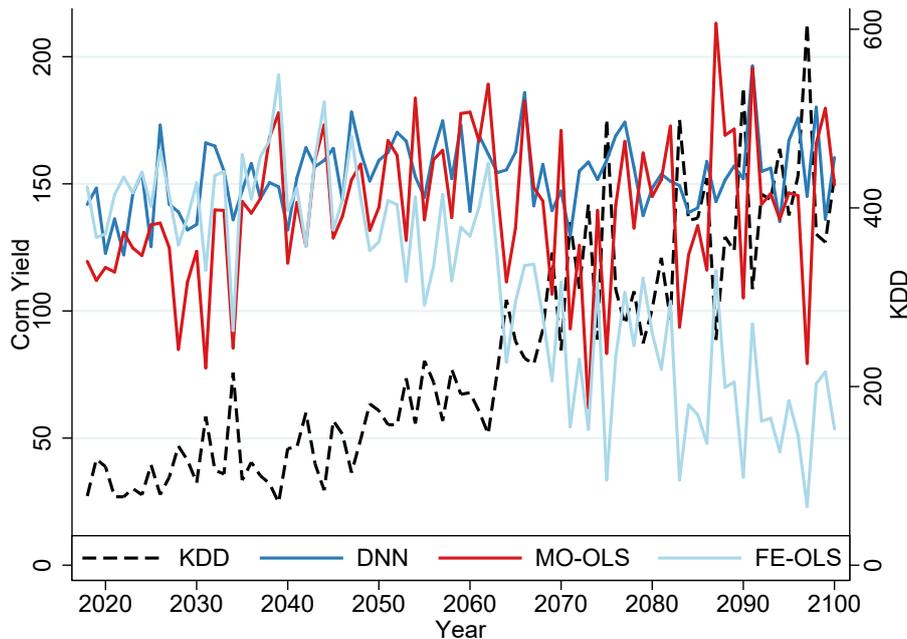
Note: This graph presents projected average corn yield using the RCP85 emissions scenario and the GFDL climate model. The results incorporate predicted technological progress in each model.

The fundamental difference between the MO-OLS model and the DNN is that the former predicts corn yields will become much more volatile due to the increased year-to-year volatility of KDD. The DNN smooths out yield fluctuations due to weather (just as it did in the 2006-15 holdout). Again, we view the MO-OLS predictions as more plausible, as it predicts yield fluctuations much more accurately in the 2006-15 holdout.

Finally, it is worth commenting on the plausibility of these predictions of the impact of technology. The corn plant is not especially sensitive to high temperature in the vegeta-

tive period (April-June) when most growth is under ground (see e.g. Tardieu et al. 2018). It is very sensitive to high temperature in the post-pollination 'corn filling' period, when high temperatures greatly increase the plant's water needs. In principle this might be addressed by increased irrigation, where there is room for progress as the large majority of US corn is not currently irrigated. But Marshall et al. (2015) find the scope for increased irrigation is limited because climate change will deplete water supplies. Another possibility is the continued development of more heat resistant hybrids, where there is history of success going back to the 1930s.[28] The greatest technical hurdle for corn may be that the pollination process that typically occurs in early July requires mild temperatures. Irrigation cannot solve the problem of pollen death or infertility due to high temperatures, and the genetic basis of pollen resilience to heat is still little understood.[29]

5. CONCLUSION

In his classic article that introduced multiple regression analysis into agricultural economics, Wallace (1920) analyzed effects of temperature and precipitation in June, July and August on corn yields. He stressed the importance of county-level heterogeneity in the production function. In particular, he showed how state-wide results for Iowa made little sense, as they aggregated very different processes at work in the cooler northern vs. warmer southern counties. He argued one must interpret regression results in the context of the biology of the corn life-cycle, which imply the mapping of weather to yield is nonlinear and crucially dependent on the timing of inputs. He noted that the simple regression methods available to him at the time could not accommodate this complexity.

Facing this conundrum, Wallace combined his scientific judgment with the regression results to produce tables of how he predicted different combinations of inputs at different times would affect yields in Polk county (obtaining a correlation of .92 with actual yields from 1891-1919). He concluded by arguing "For practical purposes, it is probably just as well first to get a general idea of the importance of the various factors at work by using the theory of multiple correlation, and then by applying common sense ... work out tables ... [like those] worked out in predicting the yield of corn in Polk County Iowa." To this the editor (C.F. Brooks) commented "To make a [national] study in accordance with these suggestions would probably require an impossible amount of labor for one person."

In this article we have shown how new methods (MO-OLS and neural networks) combined with the speed of modern computers, allow us to analyze US corn yields accounting for: (i) rich patterns of cross county heterogeneity, (ii) complex nonlinearities in the production process, and (iii) interactions across inputs over time. Thus, it is now feasible to implement Wallace's original vision. It is humbling to note we have not advanced beyond the clear understanding of the key econometric issues that he elucidated 100 years ago.

Our results show that the MO-OLS estimator, which allows for time and county fixed effects in both intercepts and slopes, provides a much better fit to corn yields than the traditional panel data methods that have been used in this literature. These are the two-way fixed effects estimator (FE-OLS) that allows for county and time fixed effects

---

[28]The corn plant can reduce its transpiration rate via stomatal closure during heat stress (loosely analogous to an animal going into hibernation), and there is genetic variance in this ability, which can be exploited to design heat resistant hybrids (see Tardieu et al. 2018).

[29]There is evidence that a shorter pollination time ('anthesis-silking interval') is associated with greater heat tolerance, and genetic variation in this trait has been used to breed more heat resistant strains. However, it has proven difficult to find other characteristics to exploit. One problem noted in the literature is that corn bred to be more heat tolerant tends to generate lower yields under good conditions.

in intercepts, and the mean group OLS (MG-OLS) estimator that allows for county-level heterogeneity in all parameters. The MO-OLS model fits better than these competitors both in random holdout samples generated in a Monte-Carlo cross-validation exercise, and in a holdout sample consisting of the last ten years of data (2006-15). Importantly, the MO-OLS estimates imply that the negative effect of "killer degree days" (KDD) on corn yield is less in hotter counties and time periods, implying adaptation of farming practices to high temperatures. This shows that in modeling corn yield it is important to account for both county-level heterogeneity and time effects.

We find that a deep neural net (DNN) that takes as input daily weather information can predict corn yields much more accurately than either the MO-OLS or conventional panel data models, both in-sample and in a Monte Carlo cross-validation exercise. However, the DNN performs quite poorly when we use it to predict yields in the 2006-15 holdout period. In particular, while the MO-OLS model predicts average annual yields quite accurately in the holdout sample, the DNN predicts a relatively smooth path, thus understating the sensitivity of yields to annual weather fluctuations.

Finally, we used the full set of models to generate forecasts of corn yield through 2100 based on the NOAAs GFDL-CM3 climate model under the RCP85 ("business as usual") emissions scenario. The traditional econometric models predict catastrophic drops in yield of about 60%, to levels last seen in the 1950s. In contrast, the DNN and MO-OLS models predict that corn yields will stagnate, but not actually decline. The MO-OLS model predicts less severe consequences of climate change because: (i) it predicts significant adaptation of farming practices to hot conditions, and (ii) it predicts technical progress will have a larger positive effect on yields.

A key difference bewteen the DNN and MO-OLS is that the DNN generates its own forecast of future technical change, while we must input a forecast into the MO-OLS model. This is not necessarily an advantage for the DNN, as its forecast is a "black box." Thus, we can't use the DNN to do scenario evaluations for different degrees of adaptation or technical progress. And, in contrast to the MO-OLS model, we can't really decompose why the DNN expects less impact of climate change than traditional econometric models into parts due to adaptation or other types of technical change.

The GFDL-CM3 climate model under the RCP85 emissions scenario predicts not only that the average annual level of $KDDs$ will increase substantially in the second half of the 21st century, but also that the year-to-year volatility of KDD will increase substantially. Our MO-OLS model predicts that this will lead to a dramatic increase in year-to-year volatility in yields. In contrast, the DNN predicts much less sensitivity of annual corn yield to KDD, and hence it predicts a much smoother yield path. We view the MO-OLS model prediction as more credible, as it accurately predicts year-to-year yield fluctuations in the 2006-15 holdout sample, while the DNN predicts a yield path that is far too smooth.

Substantively, it is important to note even stagnation of US corn yields is likely to create very serious or even catastrophic problems for future world food supply. While here we exclusively look at the RCP85 scenario, Keane and Neal (2020) also examine the RCP45 and RCP26 scenarios and find that emissions reductions akin to those proposed in the Paris agreement could greatly mitigate the damages to yield found here.

Finally, a key takeaway from our analysis is the impressive performance of the MO-OLS estimator. It is significant that MO-OLS generates both (i) more accurate forecasts of yield than the traditional panel data models (both in-sample and out-of-sample), and (ii) forecasts for the 2016-15 holdout that are much more accurate than the DNN, while using many fewer parameters and maintaining the interpretability of the estimates.

REFERENCES

Burke, M. and K. Emerick (2016). Adaptation to climate change: Evidence from US agriculture. *American Economic Journal: Economic Policy 8*(3), 106–140.

Butler, E. and P. Huybers (2013). Adaptation of US maize to temperature variations. *Nature Climate Change 3*, 68–72.

Clevert, D., T. Unterthiner, and S. Hochreiter (2015). Fast and accurate deep network learning by exponential linear units (elus). *CoRR abs/1511.07289*.

Frisch, R. and F. Waugh (1933). Partial time regressions as compared with individual trends. *Econometrica 1*(4), 387–401.

He, K., X. Zhang, S. Ren, and J. Sun (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR abs/1502.01852*.

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2012). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR abs/1207.0580*.

Ioffe, S. and C. Szegedy (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR abs/1502.03167*.

Kawasaki, K. and S. Uchida (2016). Quality matters more than quantity: asymmetric temperature effects on crop yield and quality grade. *American Journal of Agricultural Economics 98*(4), 1195–1209.

Keane, M. and T. Neal (2020). Climate change and u.s. agriculture: Accounting for multidimensional slope heterogeneity in production functions. *UNSW Research Working Paper 2018-08A*. Available online: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3180480`.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *CoRR abs/1412.6980*.

LeCun, Y., L. Bottou, G. Orr, and K. Müller (2012). Efficient backprop. In G. Montavon, G. Orr, and K. Müller (Eds.), *Neural Networks: Tricks of the Trade*, Volume 7700 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg.

Leshno, M., V. Ya. Lin, A. Pinkus, and S. Schocken (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks 6*, 861–867.

Li, X., S. Chen, X. Hu, and J. Yang (2018). Understanding the disharmony between dropout and batch normalization by variance shift. *CoRR abs/1801.05134*.

Lobell, D., M. Banziger, C. Magorokosho, and B. Vivek (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Climate Change 1*, 42–45.

Lobell, D., G. Hammer, G. McLean, C. Messina, M. Roberts, and W. Schlenker (2013). The critical role of extreme heat for maize production in the united states. *Nature Climate Change 3*, 497–501.

Marshall, E., M. Aillery, S. Malcolm, and R. Williams (2015, November). Climate change, water scarcity, and adaptation in the u.s. fieldcrop sector. Technical Report Economic Research Report 201, United States Department of Agriculture.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Reclamation (2013). Downscaled cmip3 and cmip5 climate projections: Release of downscaled cmip5 climate projections, comparison with preceding information, and summary of user needs. U.S. Department of the Interior, Bureau of Reclamation, Technical Service Center, Denver, Colorado, 116 p., available at: `http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/downscaled_climate.pdf`.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *CoRR abs/1609.04747*.

Schlenker, W. and M. Roberts (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences 106*(37), 15594–15598.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*, 1929–1958.

Tannura, M. A., S. H. Irwin, and D. L. Good (2008). Weather, technology, and corn and soybean yields in the u.s. corn belt. *Marketing and Outlook Research Report 2008-01*.

Tardieu, F., T. Simonneau, and B. Muller (2018). The physiological basis of drought tolerance in crop plants: A scenario-dependent probabilistic approach. *Annual Review of Plant Biology 69*, 733–59. Available online: `http://www.farmdoc.uiuc.edu/marketing/morr/morr_archive.html`.

Thompson, L. (1988). Effects of changes in climate and weather variability on the yields of corn and soybeans. *Journal of Production Agriculture 1*, 20–27.

Wallace, H. A. (1920). Mathematical inquiry into the effect of weather on corn yield in the eight corn belt states. *Monthly Weather Review 48*, 439–446.