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 The econometric literature on predicting impacts of climate change on crop yields has produced mixed results. The problem is difficult, because the mapping from weather to crop yield is high dimensional and nonlinear. We consider the potential for deep neural networks (DNNs) to generate better predictions. Using U.S. county-level corn yield data from 1950-2015, we show that DNNs offer substantial improvements over standard econometric methods for predicting yields, both in-sample and out-of-sample. However, the mechanisms underlying DNN predictions are a 'black box.' Thus, we also report results from a new panel-data estimator developed by Neal (2018) called MO-OLS, that allows for fixed-effects in intercept and slope coefficients when estimating the agricultural production function. MO-OLS achieves forecasting performance close to DNNs, while maintaining ease of interpretability of the parameters. Finally, we compare predictions of traditional panel-data estimators, DNNs and MO-OLS for the impact of climate change on corn yield from the present to 2100.

Keywords: *Machine Learning, Productivity, Panel Data, Climate Science.*

1. INTRODUCTION

It is difficult to name a more important public policy issue than the potential impact of climate change on agricultural yields. In a worst case scenario the effects on world food supply could be catastrophic, yet the existing econometric literature that attempts to predict the impact of climate change on crop yields has produced mixed results. This is not surprising, given the inherent difficulty of the exercise. Unfortunately, this means the social and economic costs of future climate change are highly uncertain.

This paper considers the potential of deep neural networks (DNNs) to address two fundamental challenges in predicting the impact of climate change: First, the mapping from daily weather to annual crop yield is a very high-dimensional function and also highly nonlinear. Second, the long forecast horizon needed to predict the long-run impacts of climate change necessitates a model with very good out-of-sample prediction properties.

To appreciate the magnitude of the problem, note that crop yields are not determined by simple averages of temperature and precipitation over the growing season. Rather, the precise timing of shocks to daily temperature and precipitation matters, generating a system with thousands of inputs that interact in complex ways. DNNs are meant to be good at precisely these types of problems.

The first goal of this paper is to compare the performance of DNNs with conventional econometric methods for predicting U.S. county-level corn yield using detailed daily weather data from 1950 to 2015. The conventional econometric approach, exemplified by Lobell et al. (2011) and Burke and Emerick (2016), involves summarizing conditions over an entire growing season with a small set of variables. In particular, daily or hourly temperatures are aggregated into 'growing degree days' (GDD) and 'killing degree days' (KDD), which correspond to temperatures that are considered good or bad for the plant. 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.[1]

In contrast to this parsimonious approach, DNNs input a huge array of daily weather measures for the entire growing season, along with county and year indicators, and use these to nonparametrically estimate very high-dimensional nonlinear mappings from daily weather to final yields. In an extensive cross-validation study, we show that DNNs offer vast improvements in predictive power over standard econometric methods, both in terms of in-sample and, much more importantly, out-of-sample fit. Indeed, DNNs fit corn yields in-sample almost perfectly, and fit out-of-sample yields with mean squared errors (MSEs) that are lower by a factor of five compared to the conventional approaches.

However, an important limitation of DNNs is that they are indeed a 'black box.' The mechanisms driving their predictions are very difficult for an analyst to understand and interpret. This makes it very difficult for a policy maker to assess the face validity of the predictions, or to use the results to predict how changing inputs might change outcomes.

The second main goal of the paper is to report results from a new panel data estimator developed by Neal (2018), called Mean Observation OLS (MO-OLS), that allows for a very rich pattern of heterogeneity in both intercept and slope parameters when estimating the agricultural production function. This is the first panel data estimator to allow for multidimensional fixed-effects heterogeneity in slopes in large panels (a problem that has heretofore been viewed as intractable). We demonstrate that MO-OLS achieves a great improvement in forecasting performance over standard panel data approaches, yet in contrast to the DNN it also maintains ease of interpretability of the parameters.

Third, we propose an approach to interpreting DNN results by entering perturbed paths of daily max/min temperatures into the network. This analysis shows how increases in daily max/min temperature have very different effects depending on when they occur during the year. The DNN suggests that corn is most sensitive to heat in July and August, consistent with the scientific literature on plant growth.

Lastly, 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 CO2 emissions scenario (which can loosely be described as the 'business as usual' scenario).[2]

In order to predict future yields, all econometric approaches require us to predict technical progress (by forecasting future values of time-varying parameters). An interesting feature of the DNN is that it automatically forms its own prediction of technical progress in agriculture, simply because calendar time is an input variable. This is in some sense an advantage, as we don't have to predict technical change ourselves, but also a disadvan-

[1] Schlenker and Roberts (2009) and Kawasaki and Shinsuke (2016) use a similar approach but adopt a finer set of temperature bands.

[2] In Keane and Neal (2018) 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.

tage, 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.[3]

Historically, U.S. corn yields more than doubled between the 1950s and the present. The DNN predicts the growth in yields will soon stagnate, with yields in 2100 only about 15% higher than those today. It is difficult to ascertain, however, how much of this slowdown is due to climate change, a slowdown in technical progress, or other factors.

In order to predict technical change in the MO-OLS model, we fit VARs to the time-varying production function parameters. We present results from two VAR specifications that generate relatively 'optimistic' and 'realistic' technology scenarios. MO-OLS with the realistic scenario predicts a stagnation in yield growth, similar to the DNN prediction.[4]

In contrast, the traditional econometric models predict catastrophic drops in yield of about 60% by 2100 (back to 1950s levels) even under the relatively optimistic technology scenario. Which of these very different forecasts is more credible? The DNN and MO-OLS models provide much better out-of-sample fit than the traditional models in our cross-validation exercises, and the traditional models clearly underestimate yields after hot years. Given this, we view stagnation of yields (predicted by DNN and MO-OLS) as the more plausible scenario. Of course, given rapid world population growth, even a stagnation of crop yields will have serious adverse consequences for world food supply.

In summary, from an econometric perspective, it is impressive that the new MO-OLS estimator, which allows for multidimensional fixed-effects in slopes, fits the data far better than traditional panel data estimators. The MO-OLS model achieves forecasting performance almost as good as the far more heavily parameterized DNNs (based on out-of-sample fit), while maintaining ease of interpretability of the parameters. Clearly it is an important new econometric tool, particularly for very 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 (such as seed choice, the timing of planting, the use of fertilizers and pesticides, irrigation, etc.), and traditional factors of production like capital and labor. We can state that:

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

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.

Let $\mathbf{X}_{it} = (i, t, \mathbf{T}_{it}, \mathbf{P}_{it})$ denote the vector of all available data. 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 - using DDs to predict yield.

[3] 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.

[4] Under the optimistic scenario, the MO-OLS model predicts a modest growth in yield of roughly 60% from the present to 2100, which is still a substantial slowdown from growth over the past 70 years.

In addition, most recent studies include county and year fixed-effects to control for other unmeasured factors that affect yield, such as soil quality, technology, and trends in farming practices.[5,6] For instance, Schlenker and Roberts (2009) and Kawasaki and Shinsuke (2016) estimate regressions of the form:

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

where $DD_{j,it}$ is the total time the crop experiences temperatures above the j°C threshold, and $PREC_{it}$ is total precipitation during the growing season. The c_i and c_t are county and time fixed effects, while ϵ_{it} captures idiosyncratic shocks.[7]

Other authors simplify the model by splitting degree days into two intervals, those above and below 29^0C, which is a critical threshold for corn.[8] Thus we have beneficial 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:[9]

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

Both (2.2) and (2.3) can be estimated easily through two-way fixed effects OLS.

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 far less consequential than a hot day at the end of July, even though both may raise KDD by the same amount.

Some recent articles introduce non-linearity into (2.3) via slope heterogeneity. Butler and Huybers (2013) add slope heterogeneity across counties using the mean group regression estimator (MG-OLS) due to Pesaran and Smith (1995), as in:

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

One key reason to expect slope heterogeneity across i is adaptation of farming practices to county climate conditions. E.g., an extra unit of KDD_{it} may have a smaller adverse

[5] 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}.

[6] 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}.

[7] Schlenker and Roberts (2009) use State specific quadratic time trends in their main model. We use year fixed-effects instead to maintain comparability with other models we consider. Their Appendix A12 shows that using year fixed effects vs. State quadratic time trends makes little difference to their results.

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

[9] Burke and Emerick (2016) compare the panel estimator of (2.3) with a long-difference equation, and specify precipitation differently.

effect on yield in counties with relatively hot climates, if farmers in such counties adopt techniques that make the plant less heat sensitive. A limitation of (2.4) however, is that it ignores the possibility that heat and precipitation sensitivity of plants may also change over time due to adaptation, as well as other changes in technology and soil quality.

Thus, Keane and Neal (2018) allow for fixed-effects slope heterogeneity across both counties and time in an attempt to capture as much variation as possible in the response of yield to temperature and precipitation:

$$y_{it} = c_i + c_t + \beta_{1,it} GDD_{it} + \beta_{2,it} KDD_{it} + \beta_{3,it} PREC_{it} + \beta_{4,it} PREC_{it}^2 + \epsilon_{it} \qquad (2.5)$$

where $\beta_{k,it} = \beta_k + \beta_{k,i} + \beta_{k,t}$. The parameters of (2.5) can be consistently estimated using the method developed in Neal (2018), called mean observation regression (MO-OLS).

The structure we place on the fixed-effects in (2.5) enables us to identify the model, as leaving the $\beta_{k,it}$ unrestricted would obviously result in more parameters than data points. Note that (2.5) contains $(5)(N+T)$ 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. 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 $\hat{\beta}_t$. Finally, construct $\hat{\beta}_{it} = \hat{\beta}_i + \hat{\beta}_t - \hat{\beta}$. This estimator is biased, but Neal (2018) shows the asymptotic bias can be calculated exactly and removed. He also provides asymptotic and Monte Carlo results. Consistency relies on large N and T, which is appropriate given our data.[10]

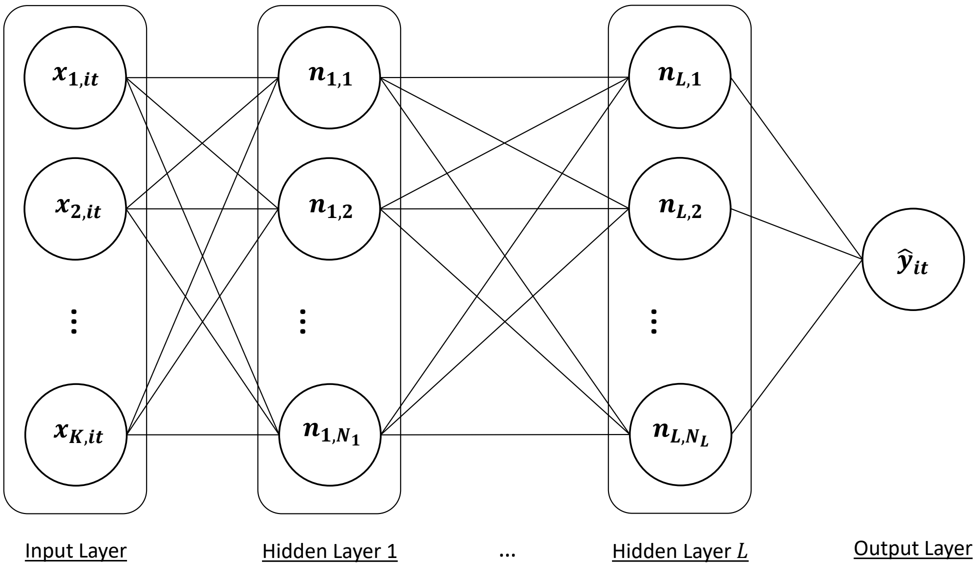
Next, we consider using a deep neural network (DNN) as an alternative to conventional econometric approaches to predicting yield. A DNN is a nonparametric nonlinear model that uses the full information set $\mathbf{X}_{it}$ (i.e., daily values for max/min temperature, daily precipitation, the year, county dummies, and finally GDD_{it} and KDD_{it}) to predict yield. This preserves the information on the timing of weather shocks and allows for any degree of non-linearity that the DNN considers to be useful to better predict yields. While it possesses many attractive features, the DNN is something 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.[11]

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 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.

[10]In theory, MO-OLS is equivalent to a 'brute force' approach where we interact all regressors with a full set of dummies for each i and t and run OLS. This would require a regressor matrix of dimension $(N + T)(K + 1)$ by NT. In our case (see Section 3.1) the number of regressors is $(N + T)(K + 1) = (2532 + 65)(4 + 1) = 12,985$, and NT is over 10^5, so it would require 17Gb of memory just to hold the regressor matrix. Thus, it would be impossible to store, let along invert, $X'X$. The MO-OLS approach 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 Neal (2018).

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

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):[12]

$$n_{1,u} = \begin{cases} e^{M_{1,u}} - 1 & \text{if } M_{1,u} < 0 \\ M_{1,u} & \text{if } M_{1,u} \geq 0 \end{cases} \tag{2.6}$$

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 (2.6) 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:

$$\hat{y}_{it} = w_{L,0} + w_{L,1}n_{L,1} + w_2 n_{L,2} + ... + w_{L,N_L}n_{L,N_L}. \tag{2.7}$$

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.

We search for the optimal weights that minimize the in-sample sum of squared errors $\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{y}_{it} - y_{it})^2$ using an extension of stochastic gradient descent called 'adaptive moment estimation' (Adam), proposed by Kingma and Ba (2014). It has become very popular in recent DNN applications for its speed and accuracy. We choose starting values for the weights using the approach of He et al. (2015). Specifically, we draw starting values for the weights from a truncated normal distribution with mean 0 and standard

[12]The ELU has some computational advantages over the RELU or sigmoid activation functions.

deviation $\sigma = \sqrt{2/I}$, where I is the number of inputs to the layer, and where any draws that are more than two standard deviations from the mean. 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.

One method we use to improve training performance is called 'batch normalization,' a popular technique introduced in Ioffe and Szegedy (2015). The inputs to neurons inside hidden layers change with 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 that is 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. This technique has proven to improve training speed and accuracy.

Another problem with DNNs is they have so many parameters they are almost guaranteed to overfit the training data. To reduce this problem, we adopt a technique called 'dropout' 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 the neighboring neurons, which helps prevent overfitting, 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 has failed to find an improvement in a set number of iterations. We implement the model using Google's *Tensorflow* package, which 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. MONTE CARLO CROSS-VALIDATION STUDY

3.1. Data Sources and Methodology

To fit the models of Section 2, we require county-level panel data on corn yield and weather. Historical temperature and precipitation data are taken from Schlenker and Roberts (2009). They contain daily observations on max/min temperature, along with precipitation, for U.S. counties from 1950 to 2015.[13] 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 2018 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) 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 $N = 2,532$ and $T = 65$.[14]

[13] The authors extended the dataset to 2015 after the publication of their paper, and provided code to map the observations across a grid to each county.
[14] Following the literature, we exclude counties west of the 100th Meridian that rely heavily on irrigation.

The complete set of models we consider is summarized in Table 1. They are classified 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' which include the raw *daily* temperature and precipitation data. All models include year and county dummies. The full information models also include the GDD and KDD variables, so as to ensure that they incorporate all available information.[15]

Table 1. Description of Models

Title	Description
	Degree Day Models
FE-OLS	Equation (2.3): OLS with county and year fixed effects.
FE-OLS w/ disagg. DDs	Equation (2.2): Same as 'FE-OLS' but uses DD intervals of 3 degrees instead of the GDD/KDD split, as in Schlenker and Roberts (2009).
MG-OLS	Equation (2.4): Mean Group OLS which allows for intercept and slope heterogeneity across counties. Includes a time trend.
Deep Neural Net	The DNN outlined in Section 2. It includes 10 layers, where the first layer features 3000 neurons and the remaining layers 2000 neurons each.
MO-OLS	Equation (2.5): Mean Observation OLS as proposed in Neal (2018). Allows for intercept and slope heterogeneity across counties and over time.
	Full Information Models
OLS w/trend	OLS which includes a year time trend.
LASSO w/trend	LASSO where the regularization parameter is selected through its own cross-validation analysis.
Deep Neural Net	The DNN outlined in Section 2. It includes 10 layers, where the first layer features 3000 neurons and the remaining layers 2000 neurons each.

[15]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.

We compare the fit of the models using the correlation between actual and fitted values, $Cov(\hat{y}_{it}, y_{it})/\sigma_{\hat{y}}\sigma_y$, and the average mean squared error (AMSE) $\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{y}_{it} - y_{it})^2$. Note that y_{it} here is the logarithm of corn yield (in bushels per acre). We present the results for both in-sample and out-of-sample fit.

To compare out-of-sample fit of the models we use Monte Carlo cross-validation. Specifically, we partition the observations prior to estimation, with 80% of the sample being used 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.

3.2. Model Fit Results

Table 2 presents fit results for all models described in Section 2, presented in the same order as in Table 1. The top three panels 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 more disaggregated temperature 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 approaches taken by (i) Burke and Emerick (2016) and Lobell et al. (2011), (ii) Schlenker and Roberts (2009) and Kawasaki and Shinsuke (2016), and (iii) Butler and Huybers (2013), respectively.[16] In each panel (i.e., for each model) the first row reports in-sample fit, while the second row reports the out-of-sample fit.

The three traditional econometric models provide remarkably similar fits to the data, and their measures of in-sample and out-of-sample fit are nearly identical. The correlations between actual and fitted values are 0.78 to 0.79, and the AMSE ranges from 0.121 to 0.124. It is interesting that neither the use of more refined temperature intervals, or allowing for county fixed-effects in the slope coefficients (via MG-OLS), produces an improvement in fit relative to the simple FE-OLS model based on GDD/KDD data.

The fourth panel of Table 2 reports results for the DNN based only on the GDD/KDD data (i.e., not utilizing the detailed daily weather data). This generates what can only be described as a stunning improvement in prediction ability over the traditional econometric models. As expected, the DNN's in-sample fit is clearly better than its out-of-sample fit, yet even its out-of-sample fit is much better than that of the econometric models. The correlation between the fitted and actual data increases to 0.92 and the AMSE drops by almost 60%, to 0.053. 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 fifth panel of Table 2 reports results using MO-OLS. The striking finding is that the out-of-sample fit of the MO-OLS model is equivalent to that of the DNN using the same data. Thus, the rich pattern of fixed-effects in both intercepts and slopes allowed by MO-OLS enables it to 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 fact that the effect of high temperatures may differ between relatively hot/mild counties and over time due to

[16]We have of course updated their approaches to render all the model specifications identical except for the differences noted in Table 1, and to use the best available current data.

Table 2. Monte Carlo Cross-Validation Results

Estimator	Corr. Coeff.	Average Mean Squared Error Total	< 1989	≥ 1989	Low KDD	High KDD
		Degree Day Models				
FE-OLS						
in sample	0.79	0.124	0.134	0.107	0.074	0.179
out of sample	0.79	0.124	0.135	0.107	0.074	0.180
FE-OLS w/ disagg. DDs						
in sample	0.79	0.121	0.131	0.106	0.075	0.174
out of sample	0.79	0.121	0.132	0.106	0.075	0.174
MG-OLS						
in sample	0.78	0.127	0.139	0.110	0.084	0.177
out of sample	0.78	0.127	0.139	0.109	0.084	0.177
Deep Neural Net						
in sample	0.96	0.026	0.027	0.026	0.016	0.036
out of sample	0.92	0.053	0.053	0.054	0.029	0.077
MO-OLS						
in sample	0.94	0.037	0.037	0.037	0.021	0.055
out of sample	0.92	0.052	0.052	0.052	0.029	0.078
		Full Information Models				
OLS w/trend						
in sample	0.88	0.073	0.071	0.077	0.043	0.103
out of sample	0.88	0.075	0.072	0.079	0.044	0.106
LASSO w/trend						
in sample	0.81	0.113	0.122	0.100	0.072	0.155
out of sample	0.81	0.113	0.122	0.099	0.072	0.154
Deep Neural Net						
in sample	1.00	0.003	0.003	0.003	0.002	0.004
out of sample	0.96	0.028	0.025	0.032	0.014	0.041

Note: This table contains results of the Monte Carlo cross-validation study using models and data outlined in Section 3.1 and Table 1. All results are averages across 100 Monte Carlo iterations, except the Deep Neural Net results which are averages of 10 Monte Carlo iterations.

adaptation behavior by farmers. A careful analysis of the distribution of the fixed effects reveals that this in fact occurs (see Keane and Neal (2018) for details).

Next we turn to the bottom of Table 2, which reports results for the full information models that contain well over a thousand covariates. The sixth panel of Table 2 reports OLS results using the full set of covariates. Interestingly, this provides a substantially better fit than the three traditional econometric models that summarize the temperature and precipitation data by the annual variables GDD, KDD and $PREC$. For example, the out-of-sample AMSE is 0.075, compared to 0.124 for the FE-OLS model in the first row. It is also interesting that the in-sample and out-of-sample fit statistics are nearly identical, so there is no evidence of it overfitting the data. This may seem surprising given we have over a thousand covariates, but recall that we also have $126,043$ data points.

The seventh panel of Table 2 reports LASSO regression results. As there is no evidence of overfitting for OLS, it is perhaps not surprising that LASSO leads to a deterioration in predictive ability. But it still does a bit better than the standard econometric models.

The last panel 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 correlation between predicted and actual yields, and a correlation of 0.96 between predicted and actual out-of-sample yields. The out-of-sample AMSE is only 0.026, which is half as large as that of the DNN and MO-OLS models that use only the annual summary statistics KDD, GDD and $PREC$ as inputs. Furthermore, it is four times more accurate (based on AMSE) than the traditional econometric models in the first three panels of Table 2. Thus, the DNN is 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 the models that rely on aggregate summary statistics.

Note that we do not implement MO-OLS with the daily temperature data, as it would be infeasible to estimate fixed-effects in the slope coefficients for so many covariates. However, it may be possible to estimate MO-OLS models that add additional covariates with more precise timing or contextual information. We leave this for future research.

To summarize our results so far, the DNN and MO-OLS degree day models outperform the traditional econometric models (in terms of out-of-sample AMSE) by a factor of about 2.4, while the DNN that uses the full set of daily covariates outperforms them by a factor of 4. While the MO-OLS model based on the degree day information does not fit the data as well as the DNN using full information, it retains the advantage that its results are more easily interpretable, and it can be used for scenario evaluations (while this is very difficult using the DNNs). This will become more clear in subsequent sections.

A key take away is that our results demonstrate it is possible to get fairly close to the predictive power of DNNs using a more easily interpretable econometric estimator (MO-OLS) that relies on a small fraction of the number of parameters and covariates.

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 four columns of Table 2 compare model fit: (i) before/after 1989, when Keane and Neal (2018) find a structural break in the KDD coefficient) and (ii) cases where KDD was above/below average (i.e., hot vs. cold counties/years).

Note that the three traditional panel data models all fit a bit better in the post-1989 period. Both the DNN degree day model and MO-OLS fit equally well before and after 1989. Interestingly, in the bottom rows of the table, we see that the OLS and DNN models based on the raw daily weather data fit slightly better (out-of-sample) *before* 1989.

Differences are much greater when we compare low KDD vs. high KDD cases. Every model has an AMSE for high KDD cases that is more than double that for low KDD cases. Thus, all models have greater difficult in accurately predicting yields in hot counties/years. However, note that even in high KDD cases the MO-OLS and DNN degree day models produce out-of-sample AMSEs similar to that generated by the traditional panel data models in the easier to predict low KDD cases.

Furthermore, the DNNs based on the raw daily weather data fit the low KDD counties/years almost perfectly (i.e., out-of sample AMSEs of only 0.014 to 0.015). Even though they are less accurate in high KDD situations (i.e., out-of sample AMSEs of only 0.038 to 0.040), they are still twice as accurate in these cases as the traditional models in the easier to predict low KDD cases. This highlights the truly dramatic across-the-board fit improvement that this DNN achieves.

Finally, Table 3 compares the distributions of out-of-sample prediction errors for actual yield across our main models. The top half reports unconditional results, while the bottom half focuses on the above average KDD cases. A key take away from the table is that the three traditional econometric models generate larger underestimates of yield in the the high KDD cases than in the overall holdout sample. Furthermore, their 10th percentile forecast errors in the high KDD cases are 50 to 60 bushels per acre, which is roughly a 50% underestimate. This poor performance in high KDD cases suggests the ability of the traditional econometric models to predict impacts of climate change is questionable, as climate models predict hot growing seasons will be much more common in the future.

Table 3. Prediction Errors for Yield by Estimator

	Mean	Median	10th pct.	90th pct.
All out-of-sample data				
FE-OLS	-4.73	-1.65	-39.29	27.06
FE-OLS w/ disagg. DDs	-4.69	-1.47	-39.03	26.82
MG-OLS	-4.53	-3.67	-38.54	27.25
MO-OLS	-0.81	-0.85	-20.04	18.67
Deep Neural Net (Full Info.)	1.79	1.42	-02.34	06.74
High KDD out-of-sample data				
FE-OLS	-7.32	2.71	-60.69	27.79
FE-OLS w/ disagg. DDs	-7.52	2.39	-61.16	27.19
MG-OLS	-5.90	0.54	-51.88	25.59
MO-OLS	-2.55	-1.75	-26.18	19.77
Deep Neural Net (Full Info.)	1.87	1.42	-02.78	07.47

Note: All figures refer to yield in bushels per acre (not log yield).

In sharp contrast, Table 3 shows that the DNN does not suffer from this problem. It's distribution of forecast errors is almost identical in the high KDD cases vs. the full holdout sample. MO-OLS falls in between. It exhibits a slight tendency to underestimate yield in the high KDD cases, but this tendency is much less pronounced than for the conventional models. The bottom line is the MO-OLS and DNN models are much less likely to underestimate yield in high KDD cases than conventional approaches.

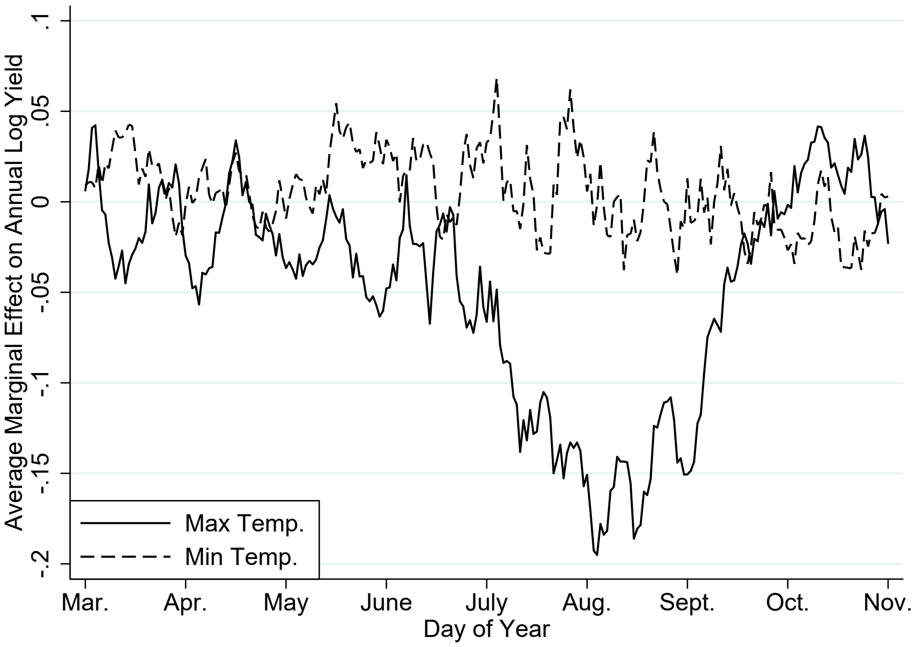
4. INTERPRETING THE NEURAL NETWORK

The Monte Carlo cross-validation results show that DNNs using raw daily weather data provide extremely good out-of-sample prediction properties. The fact the DNN using the daily data fit much better than the DNN using only annual degree data data 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%.[17] The effects are much smaller earlier and later in the growing season.[18] It is en-

[17]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.
[18]Increases in minimum temperature have much smaller effects and lack any clear pattern.

couraging that the DNN generates results that appear consistent with scientific evidence on the effect of temperature on the corn plant over the life-cycle.[19]

5. PREDICTING YIELD UNDER CLIMATE CHANGE

5.1. Methodology

Here, we compare predictions of several models presented in 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 CO2 emissions scenario. We use the NOAA's GFDL-CM3 model.[20] 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 GFDL model 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 (2018) we analyze the variability of forecasts over several different CO2 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 CO2 scenario and the NOAA's GFDL-CM3 climate model. In particular, given the dramatic out-of-sample forecasting improvements offered by MO-OLS and the DNNs over the traditional econometric models of corn yield (see Section 3), it is interesting to see if these approaches also generate meaningfully different forecasts of future yield.

As we discussed in the introduction, to forecast future yields using the econometric models we need to forecast future values of the time varying parameters that capture technical progress. For the MO-OLS model, we forecast yields using the relationship:

$$\hat{y}_{it,MO} = \hat{\delta}_{0it} + \hat{\delta}_{1it} GDD_{it} + \hat{\delta}_{2it} KDD_{it} + \hat{\beta}_{3i,t=2015} PREC_{it} + \hat{\beta}_{4i,t=2015} PREC_{it}^2 \quad (5.8)$$

Note there are three time varying parameters: the intercept, the GDD coefficient and the KDD coefficient.[21] Keane and Neal (2018) find the KDD coefficient is very well approximated by the nonlinear function $\hat{\delta}_{2,it} = 0.0025(log(KDD_{it})) - 0.0183$, obtained by regressing estimates of $\hat{\beta}_{2,it}$ from equation (2.5) on $log(KDD_{it})$. 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 (2018) also find significant variation in the intercept and GDD coefficient over time, but do not model it in a way that can be used to forecast future values. Instead, we forecast $\hat{\delta}_{0it}$ and $\hat{\delta}_{1it}$ using a VAR estimated on the historical values obtained

[19] 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.

[20] 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).

[21] We hold the coefficients on precipitation fixed because they exhibit no significant variation over time in the historical data.

from estimates of equation (2.5). Specifically, we use a VAR(1) system of two equations, which include first lags of $\hat{\delta}_{0it}$ and $\hat{\delta}_{1it}$ along with time trends.

Recall that equation (5.8) is for log yield. Thus, in specifying the time trends, we rule out using t and t^2 as that would permit exponential yield growth in levels. Instead, we consider two types of time trend: The first is $log(t)$, which we call the 'realistic' scenario of technical progress. The second includes both $log(t)$ and $(log(t))^{1/2}$, which we call the 'optimistic' scenario, because a negative coefficient on $(log(t))^{1/2}$ allows diminishing returns to technology to set in more slowly. We present results using both these scenarios of future technology, without commenting on their likelihood. To conserve on space we do not present the estimates of the VAR models.

For the two FE-OLS models (based on GDD/KDD or degree day intervals), we predict yields using (2.2) or (2.3), so we have to forecast the future values of the year effects c_t. We make these forecasts using univariate VAR(1) models fit to the historical year-effect estimates, using both types of time trend mentioned above. The MG-OLS model in (2.4) captures technical progress via a time trend, so forecasting time effects is trivial.

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 MG-OLS, 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 $\hat{\delta}_{0it} = \hat{\beta}_{0i,t=2015}$ and $\hat{\delta}_{1it} = \hat{\beta}_{1i,t=2015}$ while allowing $\hat{\delta}_{2it}$ to vary with KDD as discussed earlier.

As we discussed in the introduction, 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).[22] This contrasts sharply with the econometric models, where the time variation in intercepts and slopes has a clear technological interpretation.[23]

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

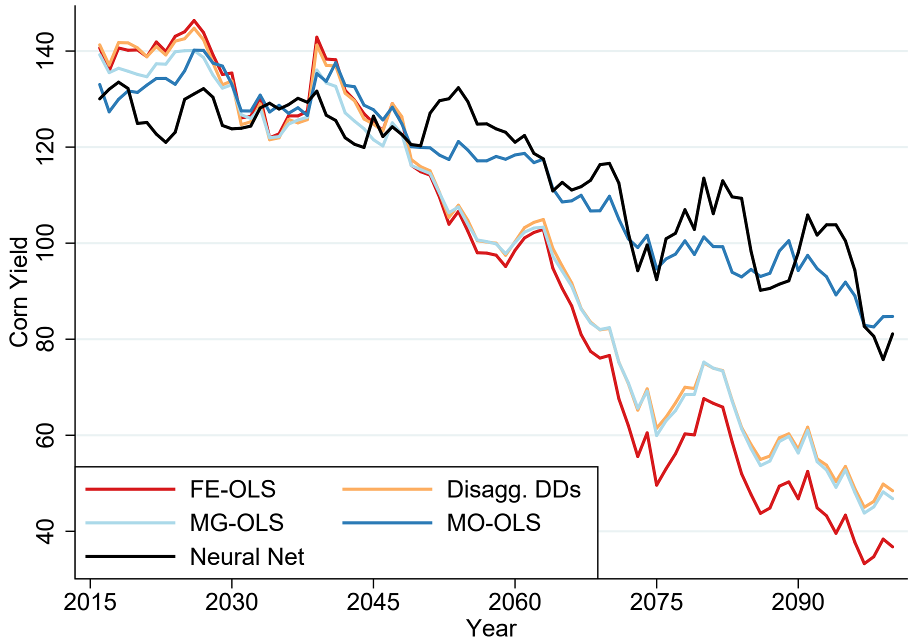
[22]Conversely, the DNN may even use trends in the other inputs (like the weather variables) to help capture technical change. The fundamental problem is that the DNN is a 'black box' and while it forecasts very well we do not fully understand the underlying mechanisms at work.

[23]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.

5.2. Future Yield Forecast Results

Figure 3 presents predictions of corn yield from 2016 to 2100 from: (i) the three traditional econometric models: FE-OLS (using either GDD/KDD or refined DD data) and MG-OLS, (ii) the MO-OLS model, and (iii) the DNN based on daily weather data. The forecasts all use weather predictions from the NOAA's GFDL model under the RCP85 emissions scenario. The forecasts in Figure 3 hold technology fixed at 2015 levels in the econometric models, and hold time fixed at 2015 in the DNN, so as to attempt to gauge the impact of climate change alone in the absence of technical progress.

Figure 3. Prediction of Corn Yield under Climate Change with No Technological Progress



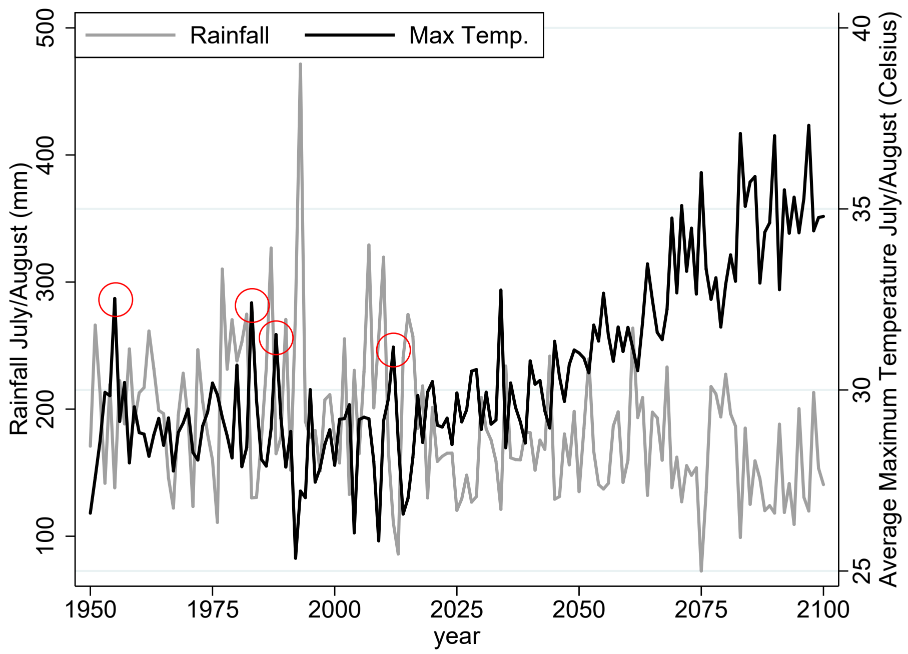
Note: This graph presents predicted average corn yield using the RCP85 emissions scenario and the GFDL climate model. The results are presented as a weighted average across U.S. counties, and time is held fixed at 2015 in each model.

The three traditional econometric models predict very similar catastrophic consequences of climate change. They predict that corn yields will drop by roughly 2/3 by 2100, bringing them back to 1950s levels. Interestingly, with the technical change (time effects) shut down, 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 to levels last seen in the 1970s and 80s.

To put these figures in some context, Figure 4 plots average rainfall and maximum temperature in July/August for the corn-growing counties of Iowa (the largest corn producing State) from 1950 to 2100, by combining the historical data and the GFDL-CM3 predictions under the RCP85 scenario. Over the historical period, there were four years (marked by red circles) in which the average maximum temperature was over 31^oC. These correspond to years of extreme drought: 1955, 1983, 1988, and 2012. The last significant drought year in the US was 2012. In that year, temperatures in July/August were unusually high and rainfall historically low.[24] As a result, corn yields came in 26% below the USDA's early (March) predictions (which were based on simple trend projection).

Figure 4 shows that the GFDL-CM3 climate model predicts conditions at least as bad as the summer of 2012 will become the norm, rather than the exception, from around 2060, and in the 2090s the average maximum will be 35^oC. Given this, it is not surprising that the MO-OLS and DNN models predict that, with no substantial effort to abate global warming and with no technical progress, typical yields will drop by 40% by 2100, as a typical year will be significantly worse than the drought of 2012.

Figure 4. Historical and Projected Weather for Iowa's corn-growing counties



Note: This figure plots the average rainfall and daily maximum temperature across July and August for all of the corn-growing counties of Iowa from 1950 to 2015. The forecast values from 2016 to 2100 are obtained from the GFDL-CM3 climate model developed for CMIP5. The red circles illustrate the four periods of extreme drought over the historical period.

Recall from Table 3 that the traditional econometric models generate seriously downward biased yield predictions under high KDD scenarios for holdout samples in the historical period, while the MO-OLS and DNN models do not suffer from this problem. This, combined with the evidence from the 2012 drought, suggests that the traditional econometric models exaggerate yield declines due to climate change (under current technology), while the MO-OLS and DNN predictions have greater face validity.

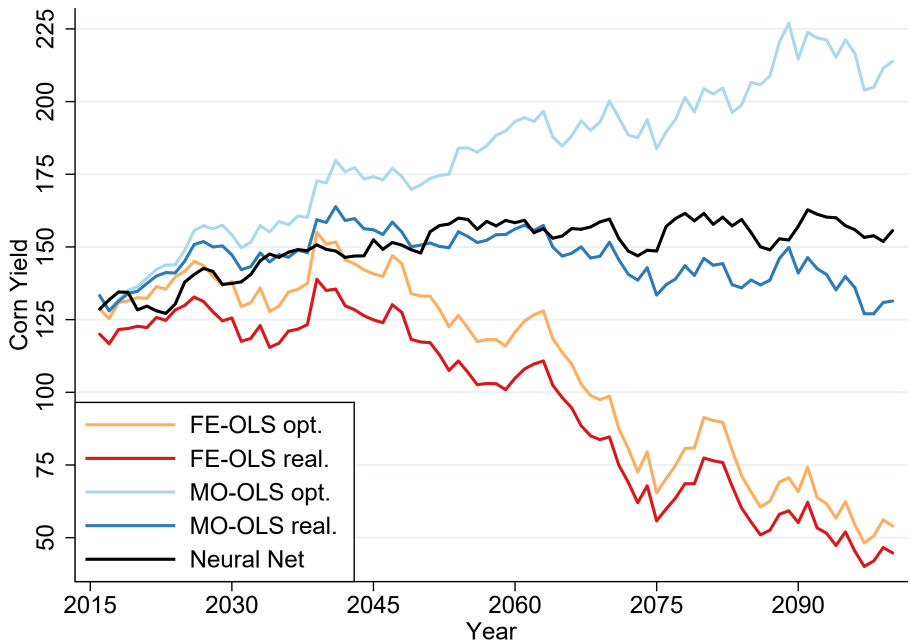
Figure 5 reports results that account for technical progress. As the three traditional

[24]Across all corn growing States, July temperatures were 5.7 degrees Farenheit above average in 2012, while precipitation was only 1.84 inches compared to an average of 3.87 inches.

economic models generate very similar results, we report only the FE-OLS model based on degree days to avoid cluttering the graph. We report results for both the optimistic and realistic technology scenarios. The striking result is that the FE-OLS model implies technical progress will do very little to mitigate catastrophic drops in yields. As we saw in Figure 3, without technical progress, the model predicts that yields will fall below 40 bushels per acre by 2100. Including technical progress only increases this to a bit above or below 50 bushels per acre, depending on whether we use the optimistic or realistic scenarios. Either way, yield is still predicted to fall to 1950s levels.

In contrast, the DNN predicts that technical progress will largely counteract the effects of 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 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 MO-OLS model under the 'realistic' technology scenario generates yield predictions similar to the DNN. It is is bit more optimistic about the next two decades, but a bit more pessimistic in the very long run. Under the 'optimistic' technology scenario, the MO-OLS model predicts that corn yields will continue to grow at a modest pace, so that by 2100 they are roughly 60% above current levels. This is still a dramatic slowdown in growth from the 1950-present period.

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 vegetative 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, but this is because high temperatures greatly increase the plant's water needs. In principle this can be addressed by (i) increased irrigation, where there is room for progress as the large majority of US corn is not currently irrigated, and (ii) the continued development of more heat resistant hybrids, where there is history of success going back to the 1930s.[25] Thus, if summer temperatures do reach the high levels predicted in Figure 4, 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.[26] Progress in this key scientific area will likely be critical for sustaining future US corn yields.

6. CONCLUSION

In his classic article that first introduced regression analysis into agricultural economics, Wallace (1920) analyzed the 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 fact, he showed that state wide results for Iowa made little sense, because they aggregated very different processes at work in the cooler northern counties relative to the warmer southern counties. He argued it was necessary to interpret regression results in the context of the biology of the corn life-cycle, which indicted that the mapping of weather to yield must be highly 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 the speed of modern computers, combined with new methods (MO-OLS and deep neural networks) allow us to analyze US corn yields accounting for: (i) extremely rich patterns of cross county heterogeneity, (ii) complex nonlinearities in the production process, and (iii) in the case of the neural net, complex interactions across inputs over time. Thus, these new methods make it feasible to implement Wallace's original vision. It is humbling to note that 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 multidimensional (time and county) fixed effects in both intercepts and slopes, provides a much better fit to corn

[25]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 substantial genetic variance in this ability, which can be exploited to design heat resistant hybrids (see Tardieu et al. 2018).

[26]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.

yields than the traditional panel data methods that have been used in this literature (fixed effect OLS (FE-OLS) and mean group OLS (MG-OLS)). Monte-Carlo cross-validation results show that MO-OLS provides substantially better out-of-sample forecasts than FE-OLS or MG-OLS. In particular, the traditional estimators tend to seriously underestimate yields following hot summers, while MO-OLS does not suffer from this problem. In our view, this makes the MO-OLS model more reliable for predicting yields under the hotter conditions expected to prevail in the future due to climate change.

When we provide MO-OLS or the DNN with temporally aggregated data on weather conditions over the entire growing season (i.e., degree days and total precipitation) they produce almost identical fits to the data (each producing a mean square out-of-sample forecast error about 2.5 times smaller than the traditional estimators in the literature). The advantage of the DNN is that it can instead take as its inputs the daily temperature and rainfall information, and determine how the exact timing of inputs over the whole growing season affects yields. When given this temporal information, the mean square error of the DNN forecast errors is further reduced by a factor of 2, showing, as Wallace (1920) noted, that the exact timing of inputs does matter.

Finally, we used the complete set of models to generate forecasts of corn yield through 2100 based on the NOAAs GFDL climate model under the RCP85 emissions scenario which features strong emissions growth. The traditional econometric models predict catastrophic drops in yield on the order of 60% to 70% (i.e., back to levels last seen in the 1940s and 50s). We are skeptical of these forecasts, because, as we noted earlier, these models generate severely downward biased estimates of yields following hot summers even in the historical data.

In contrast, the DNN and MO-OLS models predict that the growth in corn yields will stagnate, but they do not predict that they will actually decline. The reasons that MO-OLS predicts less severe consequences of climate change are that: (i) it predicts greater adaptation of corn production processes to hot conditions than do the traditional models, and (ii) it also predicts that yields will be more favorably affected by technical progress. We hasten to add, however, that even stagnation of US corn yields is likely to create serious problems for future world food supply. While here we exclusively look at the RCP85 scenario, Keane and Neal (2018) 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.

A key difference bewteen the DNN model 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' we can't evaluate its face validity. We also can't use DNN to do scenario evaluations for different degrees of adaptation or technical progress. Thus, 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.

Finally, a key takeaway from our analysis is the impressive performance of the new MO-OLS estimator. It is significant that MO-OLS generates both (i) far more accurate forecasts of yield than the traditional econometric methods (based on or cross-validation exercise), and (ii) forecasts that are close to the accuracy of the DNN, but 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, January). 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 U. Shinsuke (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 (2018). The impact of climate change on u.s. agriculture: The roles of adaptation techniques and emissions reductions. *UNSW Research Working Paper 2018-08*. 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*.

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.

Neal, T. (2018). Multidimensional slope heterogeneity in panel data models. *UNSW Research Working Paper 2016-15A*. Available online: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2853992`.

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`.

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.