Supply response at the field-level: disentangling area and yield effects

Matthieu Stigler

This version: March 10, 2020.

Latest version available here

Abstract

Agricultural price supply response is thought to occur mainly through changes in acreage rather than through yield increase. Many studies find that yields respond weakly to prices, leading to the counter-intuitive idea that yields are insensitive to prices. In this paper, I argue that this result is likely due to the use of aggregated data: county- or state-level yields are averages, whose composition itself is affected by price changes. When area expansion is done by cultivating less fertile fields or by foregoing rotation, this composition effect reduces average yields, even if yields increase on each individual field.

To disentangle the effect of the intensive and composition effect on county-level yields, I run an analysis at the field level, constructing a dataset of satellite-predicted crop choice and yield data for corn and soybeans for close to two million fields in the US Midwest. Results indicate that the field-level yield elasticity to prices is higher than reported elsewhere, ranging from 20% to 30%. When the same analysis is done using pseudo county-level panel, results change drastically for corn, becoming even negative. These results shed light on the complex dependency between area and yield response, supporting the hypothesis that area expansion is made by mainly bringing fields with lower yields.

1 Introduction

Does agricultural production respond to higher prices by increasing yields, or only by expanding area? Further, does the type of area response, whether it occurs by using high quality fields or expanding on marginal land, change the dynamics of the yield response? Intuitively, the answer will depend on whether we are using field-level data or county aggregates. With county or state aggregates, an increase in output prices bringing into production marginal lower-productivity fields will mechanically decrease average yields. This *composition* effect suggests that analyzing yield supply based on aggregated data delivers estimates that are a mixture of individual yield response and aggregate area response. This makes interpretation of the aggregate-based analysis ambiguous: when observing changes in county yields, it is not possible to infer whether these changes are effectively due to changes in individual yields, or whether they are due to a simple change in the composition of the type of fields within a county. This lack of interpretability of aggregated data has been acknowledged in many supply response studies such as Roberts et al. (2013), Beddow and Pardey (2015) or Miao et al. (2016), but remains so far unaddressed. As a typical example, Miao et al. (2016, p. 199) find that the county-yield response to prices for soybeans is close to zero statistically, conceding however that their result "could also be indicative of the intensive [yield] and extensive margin [area] effects offsetting each other". As a result, it is difficult to draw policy-relevant implications from supply response analysis, for a same estimate can be generated under opposite scenarios.

In this study, I address the non-informativeness of aggregate-based estimates by building a dataset at the field level. I use recent developments in satellite remote sensing (Lobell et al., 2015; Jin et al., 2017) to predict corn and soybeans yields at the 30m resolution for eight states in the US Corn Belt. Combining this with a dataset of crop choice and a dataset of field boundaries, I obtain a dataset of around two million fields over ten years (2008-2017). I add local corn and soybean prices, measured from close to 2000 grain elevators, and weather controls. This unique dataset allows me to run an analysis at the field level, and to compare it to an analysis at the aggregate level by constructing pseudo county-level data. Results confirm the intuition that studies based on county-level data under-estimate the individual-level yield response, due to the composition effect arising from the entry of new fields with lower productivity.

The supply elasticity estimates sought in this paper are key to many debates in economics. Given the role of the US as a large exporter of corn and soybeans, the extent to which a bad harvest in one part of the world causes higher international prices and volatility depends ultimately on the supply elasticity in the US, see for example Lybbert et al. (2014). There is also large literature on *how to feed the world* in 2050 (Ehrlich and Harte, 2015; Godfray et al., 2010; Tilman et al., 2011), investigating whether and how agriculture can respond to the predicted increase in demand. Having a good understanding of the area and yield response is crucial in this debate to identify future bottlenecks. With the demand predicted to nearly double, will yield increases be enough, or will more land be required for agriculture? The last point has also important consequences to assess the environmental costs of agricultural expansion. Supply expansion through intensification raises concerns of higher nitrogen leaching, resulting into ground-water pollution and algal blooms (Hendricks et al., 2014a). On the other side, supply expansion through extensification can reduce biodiversity and increase greenhouse gas emission if forests are converted to cropland. The debate on the environmental impacts of the ethanol mandate in the USA provides a good example where reliable supply estimates are called for. On one hand, biofuels show a potential to reduce greenhouse gas emissions (GHGs) by replacing gasoline with corn-based ethanol. On the other hand, converting forests and grassland into corn fields to produce ethanol creates important GHGs emissions too. Searchinger et al. (2008) argued that due to the negative land-conversion effect, the corn-based ethanol was actually doubling GHGs emissions, instead of reducing them by 20% as initially thought. Searchinger et al.s results are however based on the assumption that supply responds mainly through the extensive margin; assuming a higher yield elasticity lessens their conclusion.

Despite the need for a good understanding of the complex dynamics of supply response, there is a paucity of studies investigating the interlinkages between area and yield response. Most study rely on aggregate data, and hence are only informative of a net effect, which is a mixture of yield and area response. Miao et al. (2016) use county-level data in the US from 1977–2007, and is hence the closest paper to mine. Using an instrumental variable approach, Miao et al. find a rather strong area (elasticity of 50%) and yield (23%) response for corn, whereas for soy they find a non-significative yield response, yet a strong area response (62%). Haile et al. (2016) follow a similar approach at the international level, and find altogether lower elasticities, with less differences between yield and area response. Further studies estimating the area or yield supply at international or regional levels include Roberts and Schlenker (2013); Magrini et al. (2018); Haile et al. (2014); Weersink et al. (2010). While some of these studies acknowledge that they are estimating a mixed effect of yield and area response, none of them is able to address this composition bias. Studies at the field level are more rare, and focus on the area response. Hendricks et al. (2014b) use data from the USDA crop map, and find a strong area response elasticity of 40% and 36% for corn and soybeans respectively. Interestingly, they show that there is an aggregation bias between county- and field-based analysis, although the bias they are concerned with pertains to the difference between short-term and long-term response.

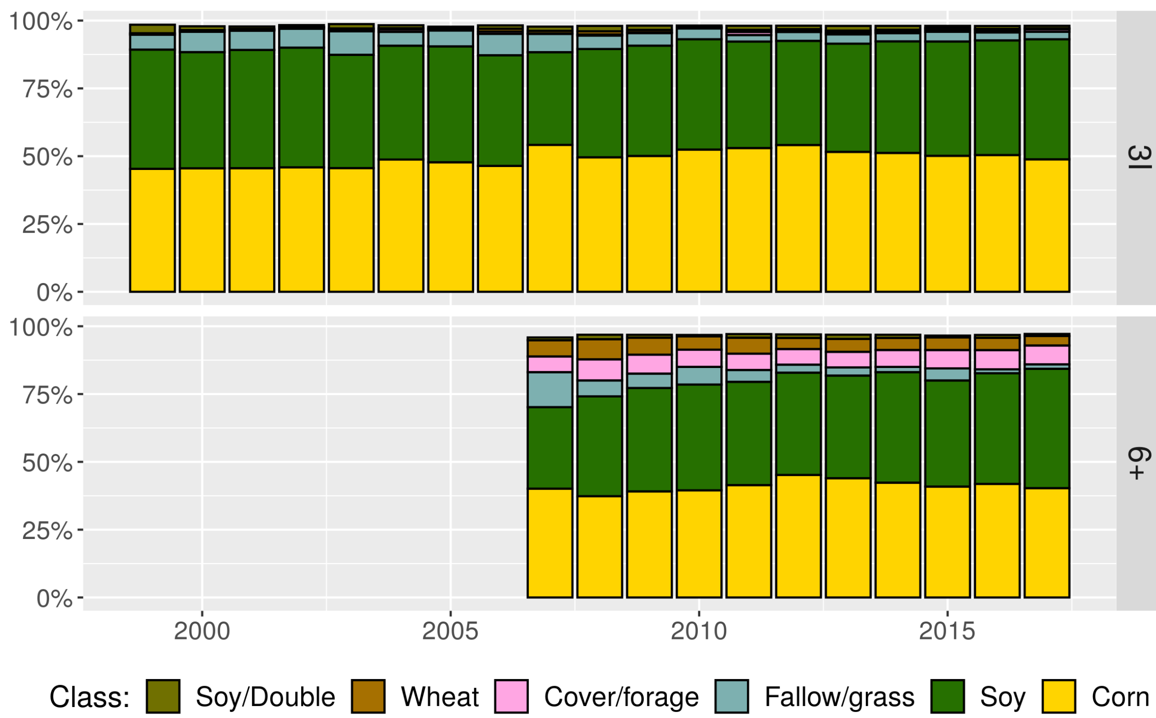
Two key elements are required for a composition effect to arise with aggregate data: there must first be a response through area expansion, and that area expansion must occur on fields that have different (typically lower) yields than the current fields. The US Corn Belt offers an interesting case in that regard, as it features two distinct types of area expansions, either by foregoing rotation or by converting marginal land. In the so-called 3I states (Iowa, Illinois and Indiana), corn and soybeans are almost the only crops, and represent more than 90% of the total cultivated area. These two crops are cultivated in rotation, which simultaneously reduces the need for fertilizer and increases yields. In these states, expansion of one of these two crops is mostly made at the expense of the other, by foregoing rotation and turning into monocropping, instead of by converting marginal land. Hendricks et al. (2014b) estimate separately area response from the rotating and marginal land, and find that the area response from marginal fields is nearly zero. This phenomenon is corroborated by various descriptive studies, see Plourde et al. (2013), Ren et al. (2016) and Stern et al. (2012), who document that the increase in the corn area at the end of the 2000-2010 decade happened mainly by foregoing rotation with soybeans in the 3I states.[1] In the other states of the Corn Belt however, the second channel of area expansion, through marginal land, is more common. Lark et al. (2015) document a net large increase in cropland of about three million acres during the 2008-2012 period. This happened mainly at the periphery of the Corn Belt, with the largest increases in South and North Dakota, followed by Southern Iowa and Northern Missouri. See also Johnston (2014) and Wright and Wimberly (2013) for similar findings.

Figure 1 illustrates the two channels of expansion, showing the shares of corn, soybeans and other crops over time. The figure distinguishes two groups of states, the 3I states as well as a second group comprising of Ohio, South Dakota, Michigan, Minnesota, Missouri, Wisconsin (referred to as *6+ states* henceforth). Three elements from the discussion above can be read from this figure: 1) the fact that the total share of corn and soybeans is much higher in the 3I than the 6+. 2) The fact that the 2007 price spike induced an expansion of corn at the expense of the soybean area. 3) The fact that there was an important conversion of marginal land into corn and soybeans in the 6+ states, at the expense of grassland and wheat (see Figure 10 in the Appendix page 32 for a detailed picture of the shares of the non corn-soy categories for each state).

What do these two expansion channels imply in terms of yield differential? For the rotational channel, the *new* fields are cropped corn after corn. There is an extensive literature (see Stigler, 2019a for a survey and refined estimates) that corn-after-corn sequences suffer a rotation penalty of about 5% to 15% for corn, and 2% to 10% for soybeans. Less is known about the yield differential in the marginal channel. Most authors conjecture that marginal land, in particular land kept under set-asides programs, is of lower quality. But on the other side, the marginal channel occurs also by

[1]This documented increase in monocroppping fields reverted back however to a larger share of fields in rotation, see Section 3.2 in the companion paper for an updated analysis up to 2017 (Stigler, 2019b).

Figure 1: Shares of crops over the 2000–2017 period



Note: Data compiled from the USDA Cropland Data Layer (CDL). Data for the 6+ group as a whole is only available 2008 onwards.

converting fields previously cropped to wheat, which are not necessarily of lower quality.

To summarize, we see that the two elements required for a composition effect are present in the Corn Belt 1) there is a rather strong area response according to most of the literature. 2) the new area seems in general to be originating from lower yielding fields. How this affects the yield response remains to be investigated. To do so, I present in Section 2 a model of aggregate yield response in presence of fields with heterogeneous quality. I begin with a marginal land model in Sub-section 2.1, and extend it in 2.2 and 2.3 to the rotation case. I describe then the dataset in Section 3 and proceed to the identification and estimation in Section 4.

2 Conceptual model

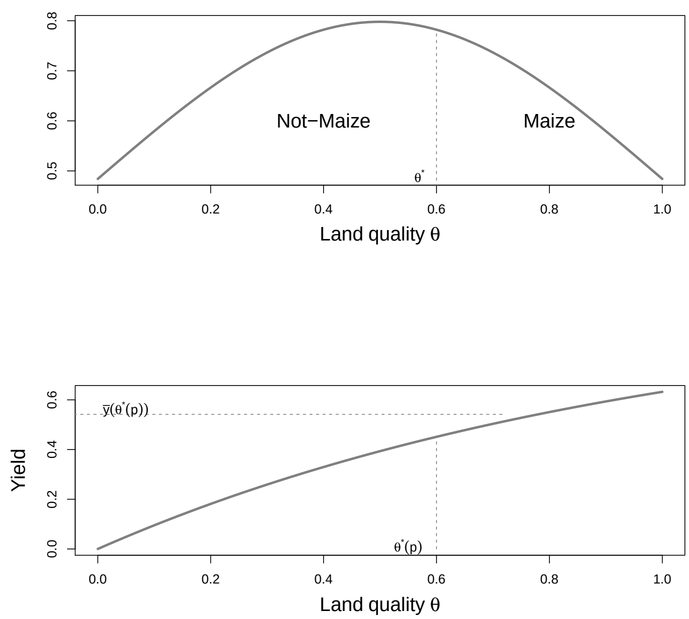
In this section, I show how the heterogeneity of fields can induce differences in supply response between aggregated or disaggregated levels. I start in Section 2.1 by building a *marginal land* model, where increases in prices induce lower fertility fields to be put into production. Given that in the Corn Belt, cropland extension is made more by foregoing rotation rather than using marginal land, I extend the model in Section 2.2 and 2.3 to the rotational margin.

2.1 Marginal land model

In his opus *On the Principles of Political Economy and Taxation,* Ricardo (1821) describes the situation where production of wheat is expanded by using less and less fertile fields. While Ricardo's interest was on how this was affecting farmer's rent, I investigate here the implications in term of aggregate supply response. Let the yield production function $y(p, \theta)$, depend on prices p and land quality θ. Field heterogeneity is modelled by introducing a field-specific land quality θ_i, with density $f(\theta)$ and cumulative function $F(\theta)$. To focus on the question of land quality, fields are assumed to be identical except for land quality; in particular, they have the same production function, i.e. $y_i = y(p, \theta_i)$. Yields are assumed to be increasing in land quality θ, i.e. $\partial y(p, \theta)/\partial\theta > 0$, as well as in prices $\partial y(p, \theta)/\partial p > 0$ as usual. Let $\theta^\star(p)$ be the *minimum quality threshold* at which profit is non-negative given a price p, i.e. $\theta^\star(p)$ is such that $\pi(\theta^\star(p), p) = 0$. Fields with higher land quality $\theta \geq \theta^\star$ enter production, while fields with lower quality ($\theta < \theta^\star$) do not produce. Normalising the land quality over the $[0, 1]$ interval, the average yield $\bar{y}(p) \equiv \int_0^1 f(\theta)y(p, \theta)d\theta$ is obtained by using the formula for the expectation of a truncated distribution:

Figure 2: Ricardian model: continuous version

(a) Distribution of θ



(b) Productivity and average yield at $\bar{p}$

$$\bar{y}(p) = \bar{y}(p, \theta^\star(p)) = \int_{\theta^\star(p)}^{1} f(\theta) y(p, \theta) d\theta / (1 - F(\theta^\star(p)))$$

Figure (2) illustrates the basic setup of the model. Panel (a) shows the distribution of land quality $f(\theta)$, and the corresponding minimum quality threshold $\theta^\star(p^1)$ for some base price p^1 above which fields are used for production. Panel (b) shows the production function over the land quality variable θ, for a given base price p^1, and the corresponding average yield $\bar{y}(p^1) = \bar{y}(p^1, \theta^\star(p^1))$.

What will be the effect of a price increase $p^2 > p^1$ on the average yield $\bar{y}(p^1, \theta^\star(p^1))$? Intuitively, we will see two effects, an increase in the *intensive* margin (higher yields) and in *the extensive* margin (more fields). The intensive effect is the direct effect of prices on yields, and is positive $\bar{y}(p^2, \theta^\star(p^1)) > \bar{y}(p^1, \theta^\star(p^1))$. The extensive effect is negative, due to our assumption of expansion through lower quality land $\frac{d\theta^\star(p)}{p} < 0$, and hence $\bar{y}(p^1, \theta^\star(p^2)) < \bar{y}(p^1, \theta^\star(p^1))$. This suggests that 1) the effect on the average is lower than the average individual effect 2) the total effect on the average yield is ambiguous, and could even be negative, despite having positive individual yields. Equation (1) formalizes this intuition (see appendix A.1 for the full derivation):

$$\frac{\partial \bar{y}(p)}{\partial p} = \frac{\int_{\theta^\star}^{1} f(\theta) \frac{\partial y(p,\theta^\star)}{\partial p} d\theta (1 - F(\theta^\star)) - f(\theta^\star(p)) \frac{d\theta^\star(p)}{dp} \left[\int_{\theta^\star}^{1} f(\theta) [y(p,\theta^\star) - y(p,\theta)] d\theta \right]}{(1 - F(\theta^\star))^2} \tag{1}$$

The first term in the numerator represents the intensive margin response, and corresponds to the *average yield response* of fields already producing. Under the standard assumption of positive individual supply response $\frac{\partial y(p,\theta)}{\partial p} > 0$, this term is positive. The second term corresponds to the composition effect of the extensive margin. It is composed of the (weighted) acreage response term $\frac{d\theta^\star(p)}{dp}$ multiplied by the average yield difference among producers $\int_{\theta^\star}^{1} f(\theta)[y(p,\theta^\star) - y(p,\theta)]d\theta$. With the assumption of yields increasing in θ ($\frac{\partial y(p,\theta)}{\partial \theta} > 0$), and of entry of lower quality land ($\frac{d\theta^\star(p)}{p} < 0$), we have that $y(p,\theta^\star) < y(p,\theta) \quad \forall \theta < \theta^\star$, so that the second term is negative.

This formula formalises the decomposition of the aggregate response into the average positive response of *incumbent* producers, and the composition effect due to the entry of new producers. As these new producers have lower yields, the composition effect reduces the aggregate response.

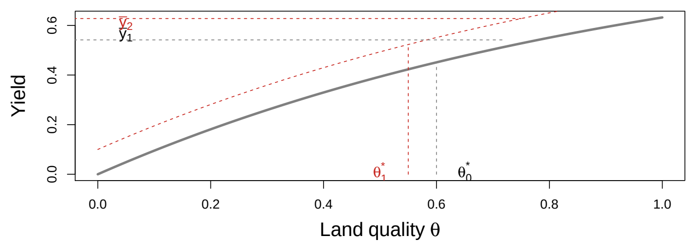
Whether the overall impact of the price increase will be negative or positive depends on the respective strength of the intensive and extensive margins. Figure 3 illustrates two possible cases. The first panel shows a case where there is a strong intensive margin (displacement of the curve) and small extensive one (new point $\theta^\star(p^2)$). The resulting average yield is higher than before. The second panel depicts the opposite situation, where a weak intensive response combined to a strong extensive response lead to a decrease in the average yield.

2.2 Modelling supply with rotation

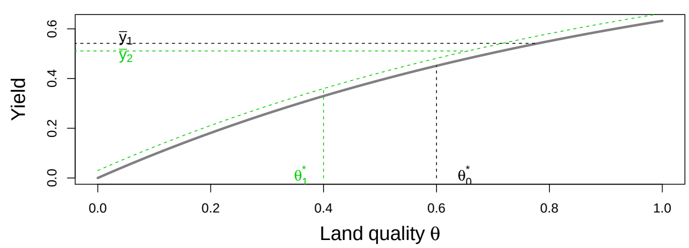
The marginal land model developed in the previous section depicts only one of the area expansion channels at play in the Corn Belt. The second channel of expansion is by foregoing rotation. I discuss here how to model individual supply with rotation, and show in the next section how to integrate rotation into the aggregate model developed in Section (2.1).

Rotation effects in production functions have been modelled in various ways in the literature, with the main differences being in the way the rotation effects are taken into account, and in the way

Figure 3: Effect of a price increase



(a) Large intensive margin, low extensive



(b) Low intensive margin, high extensive

dynamics are introduced. An early strand of literature used mathematical programming to derive optimal rules in a static framework, see El-Nazer and McCarl (1986) or Musser et al. (1985). Dynamic programming methods based on Bellman equations have been used, see Thomas (2003) on the crop choice in presence of nitrogen carry-over, or Livingston et al. (2015); MacEwan and Howitt (2011) specifically on rotations. Although interesting, these methods have the drawback that they do not lead to closed-form estimators. On the other side, the framework of Hennessy (2006) provides a clear modelling framework of rotation effects amenable to direct estimation, which I adopt here.

Hennessy (2006) considers two effects of rotation, the *input saving* effect α and the *yield boost* effect β. The input saving effect arises from nutrient carry-over from the previous period(s), and is assumed to be perfectly substituable with chemical fertiliser. This implies that the total amount of nutrient n_t available for the crop is equal to the sum of chemical fertiliser F_t and the input-saving effect α, $n_t = F_t + \alpha$. Further, this input-saving effect depends on the type of crop sequence, which we will write as α^i_j, i.e. when crop i follows crop j, leading to $n^i_t = F^i_t + \alpha^i_j$. The second effect of crop rotation is the yield *boost effect* β^i_j (for crop i following crop j), which is assumed to enter additively. These two elements lead to the following yield production function:

$$Y(F,i,j) = y^i(F_t^i + \alpha_j^i) + \beta_j^i$$

Given that crop j was planted at previous period $t-1$, crop $i^\star$ is chosen for period t if $\pi^{i^\star}(p,w,i^\star,j) > \pi^i(p,w,i,j) \forall i \neq i^\star$, where π^i is the profit function for crop i depending on the output price p and fertiliser price w. Hennessy (2006) makes the critical assumption that both the input-saving α_j^i and yield boost β_j^i effects do not depend on previous level of nutrient n_{t-1} or on actual level of fertiliser F_t. While this restrictive assumption departs from the nitrogen carry-over literature[2], it has the advantage of alleviating the need for dynamic programming tools. Furthermore, it allows us to focus on our question of interest, yield supply response in the short term.

An important implication of the perfect substitutability assumption between input saving α and chemical fertiliser is that the optimal nutrient level $n_t^\star$ does not depend on the previous crop status.[3] This in turn implies that the difference in yield for crop i between rotation $\langle ji \rangle$ or rotation $\langle ki \rangle$ is equal to the difference in respective yield boosts, i.e. $\tilde{Y}(p,w,i,j) - \tilde{Y}(p,w,i,k) = \beta_j^i - \beta_k^i$. This result is particularly important for this paper as it justifies the empirical approach relying on yields only, given that data on fertilizer use is not available (which is very difficult to infer from satellite observation).

2.3 Aggregate supply with rotational margin

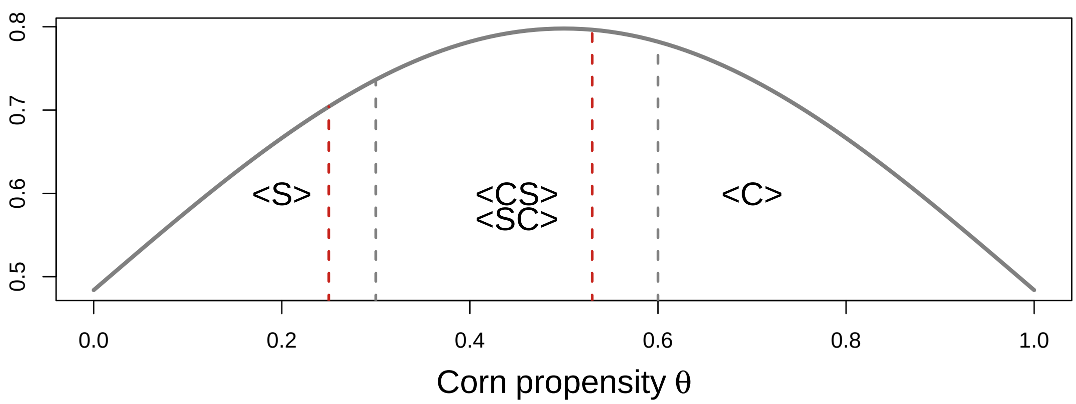
I turn now to integrate Hennessy (2006)'s model on production with rotation to the aggregate supply model developed in Section (2.1). For the sake of intuition, and in line with findings from the agronomic literature (Farmaha et al., 2016a; Porter et al., 1997), I assume that corn and soybean follow a one-year memory process, that is the rotation effect depends only on what was planted one year before. This means that we only consider four *sequences*, $C \to C$, $S \to S$, $C \to S$ and $S \to C$, which will be depicted as *rotation types* $\langle C \rangle$, $\langle S \rangle$ and $\langle CS \rangle$.

Following Hendricks et al. (2014b), I assume that that rotation types $\langle S \rangle$, $\langle C \rangle$ and $\langle CS \rangle$ are naturally ordered over a *corn-propensity* index θ. Fields with lowest propensity $\theta < \theta^1$ are cultivated to soybeans in monoculture $\langle S \rangle$, fields with intermediate propensity $\theta^1 \leq \theta < \theta^2$ are cultivated in rotation $\langle SC \rangle$, while fields with high propensity $\theta^2 \leq \theta$ are cultivated to corn in monoculture $\langle C \rangle$.

[2] Thomas (2003) uses for example a specification similar to $\alpha_j^i = m^j(n_{t-1})$.

[3] To see this, note that, for two different previous crops j or k, first order conditions $y'(F_j^{*i} + \alpha_j^i) = w/p$ and $y'(F_k^{*i} + \alpha_k^i) = w/p$ will both lead to the same available nutrient $n_t^\star = F_j^{*i} + \alpha_j^i = F_k^{*i} + \alpha_j^k$.

Figure 4: Corn-propensity model

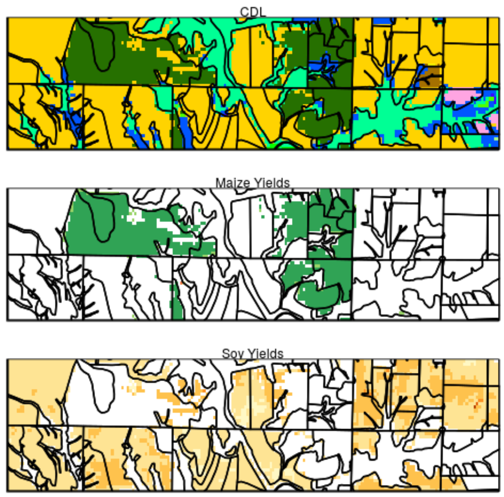


See Figure 4 for an illustration. While this hypothesis was a simple conjecture made for the sake of modelling by Hendricks et al. (2014b), I find support towards this hypothesis in related work (Stigler, 2019b), showing that fields of lower quality are most often cultivated to soybeans, whereas higher quality ones are used more often for corn.

Like in the marginal land model, a price increase will affect average yields through both the intensive and extensive margins. The difference is that while in the marginal land model we were assuming that lower quality fields were being used, here we assume that fields that are *less prone to corn monoculture* are being used. As we shall see, this raises several complications in the model. Looking first at the increase in area, we see that corn expansion following an increase in the corn to soy price occurs through a decrease in the two thresholds θ^1 and θ^2. Lowering of θ^2 corresponds to rotation fields $\langle CS \rangle$ switching now to corn-monoculture $\langle C \rangle$ fields. This is the case where the benefits of rotations are now offset by the increase in corn price, so it becomes profitable to grow corn again, even if it has a lower yield. Conversely, lowering of θ^1 corresponds to soy-monoculture fields $\langle S \rangle$ entering now a rotation scheme.

Are the new corn producing fields different from the previous ones? To answer, I assume that the production function follows the structure described in Section (2.2). That is, rotating fields $\langle SC \rangle$ will benefit from the yield boost β and fertiliser-saving α, yet the difference between rotating and non-rotating is β. In that case, the area response described above will have two opposite effects: 1) the reduction in rotating fields $\langle SC \rangle$ (which moved towards $\langle C \rangle$) reduces the number of fields with

Figure 5: Illustration of the CDL , yield and boundary data



rotation benefits, i.e. brings in fields with lower productivity, 2) the increase in rotating fields $\langle SC \rangle$ (which moved upwards from $\langle S \rangle$) brings in new corn fields with the rotation benefit, i.e. fields with higher yields.

This indicates that whether the area response brings overall fields of lower or higher yields is not as clear as in the marginal land model. Ultimately, this depends on the relative strengths of the area elasticity of the $\langle SC \rangle$ and $\langle SC \rangle$ fields, as well as on their relative shares. In practice, the share of $\langle SC \rangle$ is much more important than the share of $\langle S \rangle$: averaging over two-year sequences, 7% of the fields followed a $S \rightarrow S$ sequence, 80% a rotating one, while 15% followed a $C \rightarrow C$. This suggests that the composition effect will also be negative for rotating fields, and that we expect supply response to be lower based on average data than on individual data. The next section discusses how to estimate this in practice.

3 Crop and yield data

To conduct an analysis at the field-level, I assemble data from three main sources: a crop classification at the pixel level, a yield map for the corresponding corn and soybeans pixels, and a field boundary dataset. Figure 5 illustrates the three datasets combined. The first panel shows the crop classification, together with the field boundaries. The second panel shows the yield predictions for the pixels for which the CDL predicts maize. The third panel shows the soybeans yield predictions.

3.1 Crop classification

The crop data comes from the USDA Crop Data Layer (CDL) dataset (Boryan et al., 2011). The CDL classifies Landsat pixels of 30m$\times$ 30m into a large number of classes. The accuracy of the classification for maize and soybeans in the Corn Belt is very high, in general above 95%[4]. Corn and soybeans appear in multiple distinct classes, including categories such as corn and soybeans only, but also double-crop categories such as "Winter Wheat and Corn" or "Soybeans and Cotton". Due to the small share of the alternative classes, I focus on the main corn and soybeans class.

3.2 Yield data

The yield predictions are based on the scalable satellite-based crop yield (SCYM) method of Lobell et al. (2015) and Jin et al. (2017). The method predict yields based on a satellite-derived vegetation index.[5] Parameters linking the vegetation-index to predicted yields are derived from an agronomic crop growth model. In brief, the agronomic model is used to simulate multiple realisations of yields and vegetation index. The simulated replications are used to estimate a regression between vegetation index and yields. These estimated parameters are used in turn to predict yield based on the satellite-observed vegetation index. The advantage of this method is that it does not make use of ground data for training/calibration purpose. When ground data is available, it can be used as true validation, leading to out-of-sample (i.e. test) measure of fit, instead of in-sample measures (training).[6]

While the original papers covered 2008 to 2015, and predicted soy only for the 3I states, I extended the data till 2017, and extended soy for the five remaining states too. The accuracy of the multiple iterations of the model have been discussed in Lobell et al. (2015), Lobell and Azzari (2017) and Farmaha et al. (2016b), based either on field-level or county-level. In the data companion paper (Stigler, 2019b), I show detailed comparisons of the county-averages from SCYM against the official NASS statistics. For corn, the correlation is the highest in the three I states, at 85% at least. It is lower for the other states, with the lowest correlation being 58% for Ohio. For soybeans, results are generally less precise, the lowest correlation being 52%, again for Ohio, while the highest is 82% for Minnesota. There is ongoing work by Dado et al. (2019) to improve the model for soy, which will be integrated into this analysis when available. When validated against more than twenty thousand

[4]See `https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php`

[5]The methods uses the so-called Green Chlorophyll Vegetation index (GCVI) which is similar in spirit to the more widely known normalised difference index, NDVI.

[6]As such, the comparisons of R^2 between SCYM and direct calibrated regression in Lobell et al. (2015); Burke and Lobell (2017) are not valid as they compare training and test R^2.

corn fields, Deines et al. (2019) find that the overall correlation for corn is at 68%, ranging from 53% to 79% depending on the state. Interestingly, the model is still relatively accurate at the 30m pixel level, with an overall correlation reaching 55%.

3.3 Field boundaries

An issue with the CDL crop data is that the analysis is done at the pixel-level, while we are interested in field-level analysis. There exists however a dataset of fields boundary, the USDA Common Land Unit (CLU).[7] Unfortunately, the actual dataset is not publicly available, so that only a copy of the 2009 version can be used.

Two issues arise when using this dataset. Firstly, as the data is from 2009, fields boundaries may have changed. Drastic changes are unlikely, but cultivation of two different crops in the same field is possible. The second issue is that given that the CDL analysis is at the pixel level, instead of being at the field level, pixels in a field can contain multiple crop classes. Preliminary investigations showed clear cases of border contamination, where pixels at the edge of the field were attributed other classes (in particular classes corresponding to bush/forest elements).

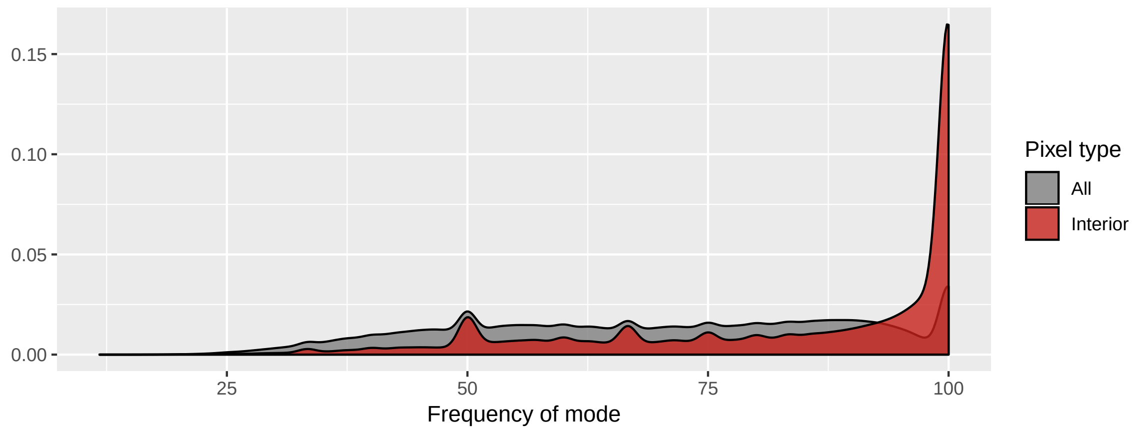
These two issues call for specific rules for the attribution of a crop to a given field. Hendricks et al. (2014b) used a *centroid-offset* rule, where the field's class is attributed according to the class of the pixel that lies at a certain distance of the field's centroid. Stevens (2015) on the other side use a majority rule, classifying the crop of a field according to the mode of its pixels. I follow a similar approach, yet make it more stringent: I set a minimum threshold on the frequency of the mode averaged over the years. The frequency of the mode can be interpreted as a measure of classification *consensus*,[8] and hence the procedure amounts to keeping only fields with high consensus. Further, I only take into account for this calculation interior pixels, i.e. pixels that do not touch the border of the field. This avoids to consider mixed pixels, that are potentially contaminated by elements outside of the field.

Figure 6 shows the frequency of the mode, with either all pixels taken into account, or only the interior ones. This is made for all plots in the eight states, using CDL classification for 2015. It is interesting to see that although the field boundaries were made in 2009, there is still a relatively good agreement for the year 2015. One can see that taking only interior pixels instead of all pixels leads to a much better result: a much larger proportion of the fields have 90% or more of the pixels showing the

[7] `https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-products/common-land-unit-clu/index`

[8] Note that classification *consensus* is not classification *accuracy*: it is possible that all pixels within a field are wrongly classified. This is however unlikely, given the relatively high classification accuracy of the CDL for corn and soybeans.

Figure 6: Agreement of pixel classification, over all fields, 2015 data



same value. One see furthermore a few bumps around the value of 50%, 66% and 75%. This suggests that the field was planted to two distinct crops (or more), using either a 1/2, 1/3 or 1/4 proportion.

To retain only fields with a good classification accuracy, the threshold was set at an average of at least 85% over all years considered (2008-2015). This is a trade-off between keepign enough fields once we want to aggregate fields over counties, and keeping well-classified fields. As a robustness checks (see 4.3), I also use a stricter version of the criterion, where I use only fields which have a *minimum* (not an average) of 85% classification accuracy over the whole sample.

3.4 Weather data

Weather variables are introduced as control variables to avoid omitted variable bias. While it is reasonable to think that weather is not influenced by prices, it is still the case that markets might anticipate weather events later in the season. To prevent this, I include a large set of weather controls from the DAYMET dataset (Thornton et al., 2017), which is at a resolution of 1000m $\times$ 1000m. The dataset includes precipitation, minimum and maximum temperatures, as well as partial pressure of water vapour. These daily measures are averaged per month, and squared terms are included. Growing degree days (GDD) will be included later on, following the work of Schlenker and Roberts (2006, 2009).

3.5 Price variables

Price variable $\boldsymbol{p}_{it}^{M}$ and $\boldsymbol{p}_{it}^{S}$ are futures quotations for post-harvest delivery (December for maize and November for soybeans), quoted pre- and in-season. The pre-planting period is defined to be the month of February and March. This is chosen earlier than actual planting times which are Mid April to May for maize, and May to June for soybeans. Given that the choice of crop is almost only between maize and soybeans, the planting period relevant to maize is also the one relevant for soybeans. Finally, this is also the period chosen by Hendricks et al. (2014b). The pre-planting price is also relevant for the yield equation, as farmers can influence yields by choosing specific types of hybrids or the sowing densities. Later on, I shall include as well a post-planting price, which shall be defined as the May-June period. This is intended to reflect within season adjustments, such as fertiliser application. Given the sunk costs already supported, it is expected that post-planting price changes will have a smaller effect compared to pre-planting ones in the yield equations.

Futures prices are adjusted for the local basis, which is taken as the difference between the closest delivery futures price and the local spot price at neighbouring elevator. The basis is measured at the same period that the price is defined, i.e. for pre-planting prices, I use an average of February-March futures (for the December maturity) and an average of the basis at the same period.

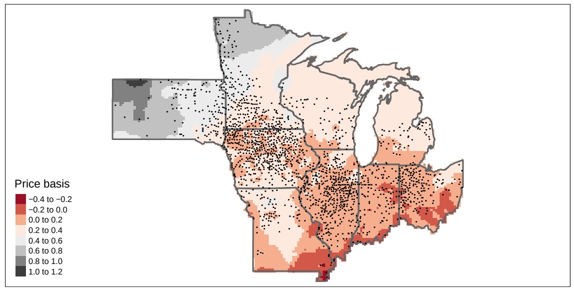
The cash prices were obtained from elevator data found in Bloomberg[9]. I end up with a dataset of close to 2000 elevators points. Data at the field level is obtained by spatial interpolation from neighbouring elevators. I use inverse distance weighting; interpolation parameters are obtained by cross-validation. It might be objected that possible transportation costs should be considered, taking for example at distance to the elevator. However, given that I use a fixed-effects strategy at the field-level, there is no need for such an adjustment, as it will get absorbed by the fixed-effects.

Figure 7 shows the location of the grain elevators and a smooth representation of the local basis. The location of the elevators follows closely where corn and soybeans are planted, compare with Figure 11 on Page 33.

On ethanol refineries There is an extensive literature (see Motamed et al., 2016 for references) finding that ethanol refineries have an impact on local maize acreage response. Motamed et al. (2016) for example find that the elasticity of maize acreage with respect to local refining capacity is about 1.5. As local refineries are likely related to the price variable, this suggests that one should add a refinery

[9]Bloomberg disseminates data originally collected by Data Transmission Network and Geograin. Data was geo-located, and databases were consolidated, averaging quotations over close vicinities.

Figure 7: Location of elevators and basis interpolation (corn, March 2014)



vicinity variable to avoid omitted variable bias. This however raises the concern that we are adding a so-called *bad control* (see Angrist and Pischke, 2008 section 3.2.3). Bad control happens when the control variable is itself endogenous to the outcome variable. This is unfortunately likely to be the case here, where location of refineries itself depends on acreage response. This is at least the argument made by Motamed et al. (2016), motivating their search for IV variables. Besides this, effects of the refinery location are likely to translate into changes in the local basis (as found by McNew and Griffith, 2005). This implies that the yield response I am measuring is also including the effect of refineries. This only changes the interpretation of the response coefficients: they include not only year-to-year variations, but also more longer term variations.

4 Identification strategy and results

The main objective of the empirical analysis is to obtain reliable estimates of the supply response at the field level. The second objective is to compare the field level estimates with estimates based on aggregate data. The research hypothesis is that yield response at the aggregated level is underestimating micro response, due to the composition effect of acreage response in average yields.

4.1 Empirical approach

The main relationship we want to estimate is the link between yields y_{it} and prices p_{it}. For corn for example, the equation is:

$$y^C_{it} = \alpha_i + \beta^C p^C_{it} + \beta^S p^S_{it} + \gamma \boldsymbol{x_{it}} + \epsilon_{it}$$

where p^C is the price of corn and p^S the price of soybeans, while $\boldsymbol{x}$ describes the set of covariates.

It should be noted that this is a highly unbalanced panel, given that farmers plant corn and soybeans alternatively.

The identification strategy relies on fixed effects at the field level, which control for time-invariant unobserved soil and farmer characteristics. There are however two main challenges in the estimation. For one, there is the threat of reverse causality, with yield response possibly affecting prices back. Second, there is a potential endogeneity between crop choice and yield response.

The issue of reverse causality from yields to prices, and that of the type of variable to proxy for price expectations, have been discussed at length in the literature. Initial supply response estimates were based on the Nerlove (1956) model, that proxies for future price expectations by using a distributed lag model of past prices. Gardner (1976) subsequently argued that using instead future prices was representing better farmers expectations. According to this argument, using futures prices for post-harvest maturity observed at pre-planting times should ensure that there is no reverse causality, since yields realized in October cannot influence prices observed 10 months before. This argument was challenged by Roberts and Schlenker (2013), who argued that markets might anticipate bad harvests, hence introducing back a source of reverse causality. Roberts and Schlenker advocated for an IV approach, using past yield shocks as instruments, while simultaneously controlling for current weather shocks. Hendricks et al. (2015) showed that the same result could be obtained without the instrument, and that one only needed to control for current weather shocks. I follow this insight in the empirical part, using pre-planting future prices and a rich set of weather controls, including monthly precipitation, temperatures and humidity, together with their squared terms.

The second identification challenge is potential endogeneity over time between crop choice and yield response. If farmers choose to grow corn only in "good" years, we will only observe a partial sample, potentially biasing the results. To address this, I make use of a unique feature of the dataset at hand. It turns out that there is a large subsample of the fields which, over the whole period considered, always rotated, or always cultivated corn.[10] The subsample of *always rotaters* is pretty large, 34%, while the fields always used for corn amount to 5%, see Figure 13 in the Appendix, page 34 for an illustration. On these fields, the area response is zero, and crop choice is perfectly predictable given last year's choice. This ensures that estimating yield response on each subsample will be free of any endogeneity between crop choice and yield.

A third estimation challenge arose, which was not initially expected. Regressions using both the

[10]There is also a subsample which always cultivated soybeans, but it is extremely small, less than 0.01%

Table 1: Regression: effect of weather controls

	Corn			Soy		
	W: none	W: early	W: all	W: none	W: early	W: all
Pre-planting Price	-0.552^{***}	-0.349^{***}	0.179^{***}	-0.230^{***}	-0.125^{***}	0.218^{***}
	(0.017)	(0.024)	(0.038)	(0.011)	(0.020)	(0.045)
Num. obs.	8297625	8297625	8297625	7596788	7596788	7596788
Num. N obs	1790910	1790910	1790910	1775658	1775658	1775658
Num. T obs (ave)	4.633	4.633	4.633	4.278	4.278	4.278
Num. variables	1	25	89	1	25	89
R^2	0.449	0.711	0.810	0.407	0.503	0.568
R^2 (proj model)	0.136	0.547	0.702	0.030	0.187	0.294

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Errors clustered at the county level.

price of corn and soybeans proved unstable to estimate, with high variations in the estimates depending on small changes in the specification. This is a typical indicator of multicollinearity, and indeed corn and soybean prices are highly correlated.[11] Unfortunately, the sample at hand starts in 2008, and does not include the 2007 high-price episode, when prices of corn and soybeans diverged substantially. To address the problem, I include only the own price in each equation, that I use the price of corn in corn equation, and soy price for the soy equation. This is justified under the assumption of the farm acting in presence of complete credit and insurance markets: once the crop is planted, say corn, the output price of the soy alternative does not play a role in the input decision for corn.

4.2 Results

I begin the analysis looking at the effect of controlling for different subsets of weather covariates. Weather covariates include monthly minimum and maximum temperatures, precipitation as well as vapour pressure, in levels and with their squared terms. I estimate three specifications, first with no weather control at all (model *W: none*), controlling then for weather in the early season pre-planting, from January to March (*W: early*) and finally controlling for the full season, from January till October (*W: all*).

Table 1 shows the regression for corn and soybeans. Coefficients are elasticities, derived from a log-log specification.

When having either no controls at all (*W: early*) or controls only for the beginning of the season (*W:*

[11] Multicollinearity measures such as the variance inflation factor were well above the value of 10 considered as evidence of multicollinearity.

Table 2: Regression: IV estimation with past weather

	Corn				Soy			
	W: none	W: none	W: all	W: all	W: none	W: none	W: all	W: all
Price OLS	-0.611^{***}		0.173^{***}		-0.243^{***}		0.314^{***}	
	(0.020)		(0.040)		(0.012)		(0.050)	
Price IV		-0.616^{***}		0.342^{***}		-0.244^{***}		0.294^{***}
		(0.021)		(0.050)		(0.013)		(0.061)
Num. obs.	7493813	7493813	7493813	7493813	6866125	6866125	6866125	6866125
Num. N obs	1783087	1783087	1783087	1783087	1765975	1765975	1765975	1765975
Num. T obs (ave)	4.203	4.203	4.203	4.203	3.888	3.888	3.888	3.888
Num. variables	1	1	89	89	1	1	89	89
R^2	0.470	0.470	0.827	0.827	0.415	0.415	0.589	0.589
R^2 (proj model)	0.160	0.160	0.726	0.725	0.031	0.031	0.319	0.319
IV Ftest		4885.279		209.648		6909.109		291.070

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Errors clustered at the county level. The sample does not contain year 2008.

early), we obtain negative coefficients. On the other side, once controlling for weather during the full season, the coefficients get a positive sign, as expected for a supply response estimate. Coefficients in the full model show an elasticity of 18% and 22% for corn and soy respectively. This is close to the estimate for corn of 23% found by Miao et al. (2016), but substantially different from their non-significant soybean estimate. These results are very interesting per se, as they substantiate the claim by Roberts and Schlenker (2013) that only using pre-planting futures price as suggested by Gardner (1976) is not enough to control for endogeneity. This is explained from the fact that there are some within-season events that can be predicted, implying that a regression on prices alone captures actually reverse causality. We see also that a rather large set of variables is needed to adequately control for weather.

Turning now to the IV estimation, I use the previous-year weather instruments suggested by Roberts and Schlenker (2013) and investigate if Hendricks et al. (2015)'s finding that they have no impact on area regressions also holds for yield regressions. Table 2 shows the IV estimates, for two of the three weather models presented above, either with no weather covariates (*W: none*) or with full-season covariates (*W: all*).

In absence of weather controls, the instruments do not have any effect, and the IV estimates are nearly identical to the OLS ones. The same happens with the model controlling for only the beginning of the season (*W: early* in Table 1), and is not reproduced here. Once controlling for weather throughout the season (*W: all),* the instruments have a noticeable effect on the corn estimate, which is now twice higher than its OLS counterpart. For soybeans on the other side, this does not happen, and the IV estimate is very close to the OLS one. These results suggest that there is for corn some elements

Table 3: Regression: subset analysis on rotating status

	Corn				Soy		
	All	Mixed	Always rotate	Always Corn	All	Mixed	Always rotate
Pre-planting Price	0.179***	0.161***	0.287***	0.202***	0.218***	0.209***	0.303***
	(0.038)	(0.039)	(0.042)	(0.051)	(0.045)	(0.049)	(0.047)
Num. obs.	8297625	5579742	2461572	256311	7596788	5184750	2412038
Num. N obs	1790910	1253629	510552	26729	1775658	1264944	510714
Num. T obs (ave)	4.633	4.451	4.821	9.589	4.278	4.099	4.723
Num. variables	89	89	89	89	89	89	89
R^2	0.810	0.813	0.806	0.709	0.568	0.557	0.584
R^2 (proj model)	0.702	0.699	0.728	0.632	0.294	0.271	0.410

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Errors clustered at the county level.

of predictability of the next harvest that is not well captured by the current weather covariates. The 2012 drought year, which particularly affected corn, could be explaining this. As a whole, the results from the IV analysis indicate that my estimates represent a conservative lower bound for the true relationship.

To investigate the second potential concern, the presence of endogeneity over time between crop choice and prices, I run the preferred model with all weather covariates, on the subsamples of fields *always rotating*, doing *always Corn*, as well as the *Mixed* ones, which have both rotating and non-rotating sequences. *Always Soy* cannot be estimated, as there are less than 0.2% of the fields following this sequence. For the sake of comparison, the full sample *All* is reproduced, showing the same coefficients as in the previous Table 1. Results are shown in Table 3. The coefficients on the *always rotating* subsample turn out to be substantially higher than the *Mixed* subsample, and higher than the full sample *All*. This suggests that there is indeed endogeneity in the crop choice, where higher prices induce an area response through marginal land and rotation-foregoing fields, both which come with lower yields, biasing the estimation of the pure intensive response that we seek here. If one focuses instead on the fields where crop choice is deterministic (either always rotating or always doing corn), the response is higher. Interestingly, we note a difference in the elasticity between fields doing always rotation and always corn, the fields doing always corn having a lower response. This is likely due to the fact that corn-monoculture fields tend to have higher yields, see Section 3.3 in the companion paper (Stigler, 2019b). Yield response to fertiliser shows strong decreasing returns (see Figure 12 in the Appendix, page 34), and mono-culture fields are more likely to lie on the flat segment of the fertilizer response curve where yields barely respond to fertiliser increases.

As a final step, I now reconstruct a pseudo county-level panel, computing county means from my

Table 4: Regression: analysis at the county level

	Corn				Soy		
	All	Mixed	Always rotate	Always Corn	All	Mixed	Always rotate
Pre-planting Price	0.001	0.001	0.118	0.141^{*}	0.196^{*}	0.206^{*}	0.178^{*}
	(0.057)	(0.057)	(0.066)	(0.068)	(0.097)	(0.097)	(0.090)
Num. obs.	6865	6864	5850	4793	6672	6669	5859
Num. N obs	704	704	598	488	692	692	598
Num. T obs (ave)	9.751	9.750	9.783	9.822	9.642	9.637	9.798
Num. variables	89	89	89	89	89	89	89
R^2	0.925	0.924	0.908	0.836	0.806	0.805	0.753
R^2 (proj model)	0.876	0.875	0.881	0.777	0.630	0.632	0.675

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Errors clustered at the state-year level.

data for each year. I do this for the whole sample, as well as for each subsample separately. Results are shown in Figure 22. We see a drastic change in the supply estimate, with the total coefficient being now even negative for corn! For corn, the result seems to be entirely driven by the *mixed* category, which is negative. Surprisingly, results for soybeans are less affected, and the estimate on the full sample is only slightly smaller than the estimate based on the field-level dataset.

4.3 Robustness checks

Clustering standard errors at different levels Table (5) the preferred model for data at the plot level is estimated using standard errors clustered at the county level. Clustering at the state level would be arguably interesting, yet is not possible owing to the fact that there are only 8 states. I use state-year clusters, which provides 80 clusters, which should be reasonable for the asymptotic approximation. For the sake of comparison I also include errors clustered at the plot level. Without surprise, errors using plot clusters are very tight, while using county or state-year errors widens them significantly, yet does not make any coefficient insignificant.

Including trend terms It is usual in yield regressions to include a time trend term, to capture among other technological increase in yields. While controlling for time seems important, it should be noted that there is a fundamental tension between controlling for time and keeping interesting time variation in our explanatory variable. Given that the price variable is comprised of the local basis and the future price, if one were to include time fixed effects, this would result into our price variable containing only the local basis, the future prices being swept out by the within-units transformation. This means that we can at most control for trend, not having a too flexible specification, as this would

Table 5: Robustness check: different clustering of errors

Crop	Subset	SE type	Estimate	P-value		Conf low	Conf high
Corn	All	Plot	0.179	< 2e-16	***	0.175	0.184
		County	0.179	1.81e-06	***	0.106	0.253
		State-Year	0.179	0.00205	**	0.065	0.294
	Mixed	Plot	0.161	< 2e-16	***	0.156	0.167
		County	0.161	4.37e-05	***	0.084	0.239
		State-Year	0.161	0.00663	**	0.045	0.278
	Always Rotate	Plot	0.287	< 2e-16	***	0.279	0.295
		County	0.287	9.50e-12	***	0.204	0.369
		State-Year	0.287	1.03e-05	***	0.159	0.414
Soy	All	Plot	0.218	< 2e-16	***	0.211	0.225
		County	0.218	1.32e-06	***	0.129	0.306
		State-Year	0.218	0.01642	*	0.040	0.395
	Mixed	Plot	0.209	< 2e-16	***	0.200	0.218
		County	0.209	2.05e-05	***	0.113	0.305
		State-Year	0.209	0.01951	*	0.034	0.384
	Always Rotate	Plot	0.303	< 2e-16	***	0.293	0.313
		County	0.303	1.26e-10	***	0.211	0.396
		State-Year	0.303	0.00217	**	0.109	0.497

have a similar effect than the time fixed effects. This is however not a large concern, as the period under consideration is rather limited, ten years, implying that a linear time trend should capture well enough the general evolution of yields. In Table 6, I show the normal specification (*none*), and compare it to a specification with a general trend (*year*) as well as state-year trends. Results are rather robust, and stay qualitatively the same.

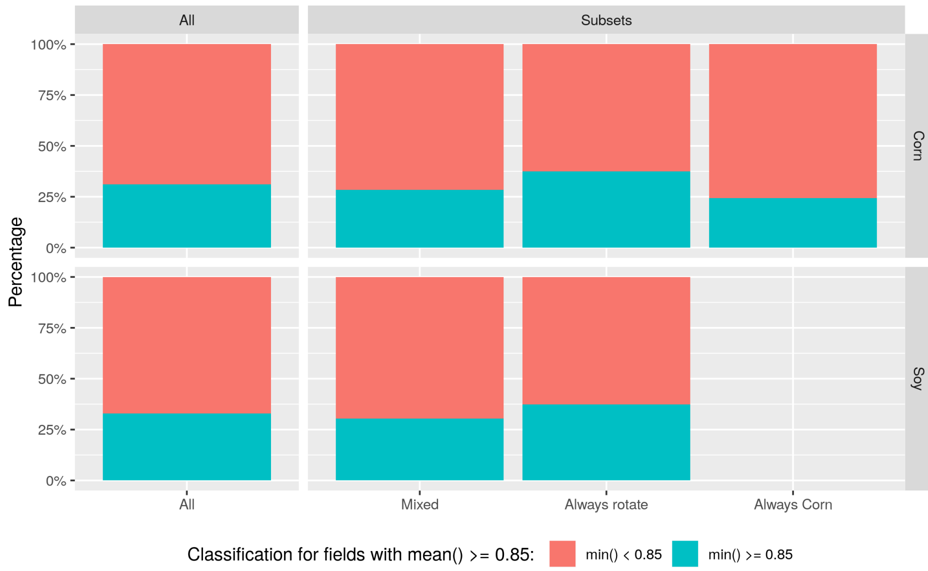
Restricting the sample to stricter classification As described in Section 3.3, tentative fields boundaries were selected for the analysis depending on the quality of the crop map classification. A field

Table 6: Regression: including time trends

	C: none	C: state-year	C: year	S: none	S: state-year	S: year
Pre-planting Price	0.179***	0.189***	0.210***	0.215***	0.158**	0.215***
	(0.038)	(0.037)	(0.037)	(0.045)	(0.050)	(0.048)
Num. obs.	8297625	8297625	8297625	7672405	7672405	7672405
Num. N obs	1790910	1790910	1790910	1783714	1783714	1783714
Num. T obs (ave)	4.633	4.633	4.633	4.301	4.301	4.301
Num. variables	89	97	90	89	97	90
R^2	0.810	0.811	0.810	0.566	0.568	0.566
R^2 (proj model)	0.702	0.704	0.703	0.290	0.292	0.290

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Errors clustered at the state-year level.

Figure 8: Field classification consensus



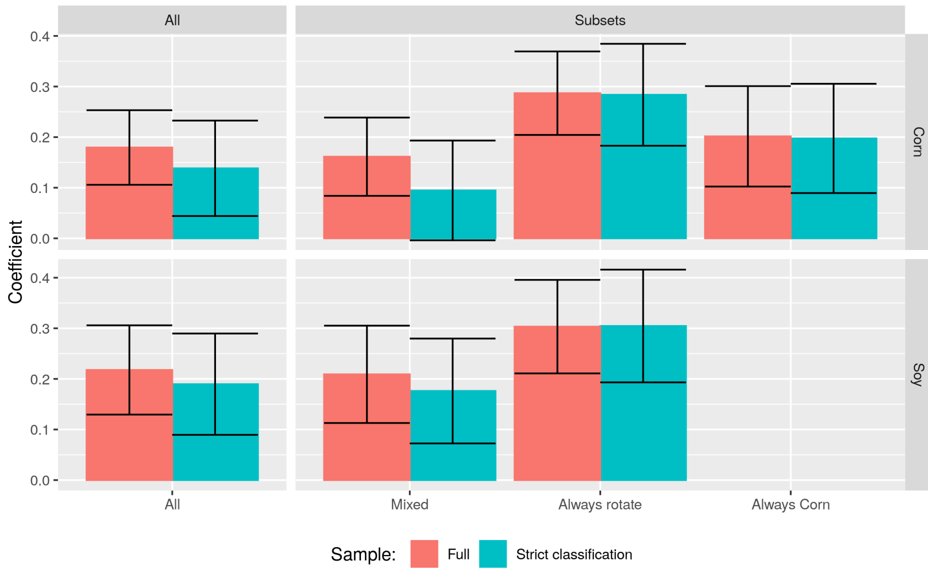
would be retained if the classification consensus[12] was at least 85% on average over the full period. This implies that in a given year, a field might be classified say to corn only at 50%, leaving the possibility that there were actually two crops on that field. If there is a strong within-field variation, this could affect our estimate of the annual yield. A much stricter criterion consists in keeping only fields whose classification consensus is at least 85% over the whole period.

Figure 8 shows the percentage of fields which have a classification consensus of at least 85%, among the fields that have on average 85% of classification consensus. The stricter classification rule keeps about 30% of the overall sample used in the analysis above. Looking at the various rotation-status subsets, the shares vary slightly: the *Always Rotating* subsample has a higher share of field with high classification consensus, for both crops.

I rerun now the analysis either on the stricter subsample of fields with at least 85% classification consensus over the whole period. Figure 9 shows the coefficients from the *Strict Classification*, and shows as well for comparison the standard coefficients using the *Full* sample. The impact is nearly negligible for the *Always Rotate* and *Always Corn* subsamples. On the other side, there is a rather large impact for the *Mixed* subsample, which sees a decrease in the elasticity, in particular for corn. The confidence intervals however always overlap, suggesting that the difference are not significant.

[12]Classification consensus is the frequency of the mode.

Figure 9: Regression: using stricter subsample



5 Conclusion

This paper investigates the yield price elasticity of corn and soybeans in the US Corn Belt. Previous literature based on county-level data found a rather small response, leading some authors to argue that yields barely respond to prices. This comes at odd with micro evidence that farmers adjust their fertiliser level or planting seed density to output prices. In this paper, I argue that this is mainly due to a composition effect owing to the reliance on county-aggregated averages. If area responds also to prices, the composition of the average changes. As area expansion is mainly done through using marginal land of lower fertility, or foregoing rotation and its benefits, the composition effect possibly reduces the average yields.

To investigate this, I build a large dataset of close to two million fields in the US Corn Belt, using a novel dataset of corn and soybean predictions based on satellite data. I find a rather high yield elasticity, of 18% for corn, and 22% for soybeans. To address concerns of reverse causality from yields to prices, I use past weather instruments as in Roberts and Schlenker (2013), and find that these only have an effect once one controls for in-season weather covariates. The IV for corn suggests that the elasticity for corn is even higher, at 30%. A second concern with the fixed effects approach here is that crop choice might be endogenous to prices. To address this, I take advantage of a specific feature of the dataset, the presence of a large number of fields who always rotate. This provides me a subsample that is free of possible endogeneity, and free of the composition effect. Estimating price yield response on this subsample, I find that supply response is higher for fields who always rotate, while it is sensibly lower for fields that changed cropping sequences over time, and hence are subject to endogeneity. I interpret this result as evidence that my estimates represent a causally identified relationship for at least a subsample of my dataset. I proceed in turn to a pseudo-county analysis, where I aggregate my fields at the county level, then run a county-panel regression, mimicking the approach followed by current literature. I find that yield estimates are drastically reduced for corn. In particular, the subsample of mixed fields that change cropping patterns over time now have a close-to-zero effect. This suggests the presence of a composition effect that biases downwards supply estimates based on county data.

The analysis presented here faces several limitations. First of all, there is non-random measurement error in the yield predictions. Soybeans in particular is less accurately estimated, possibly explaining some of the variability in the estimates found above. A bias analysis (see companion paper Stigler 2019b) shows that the direction of the measurement error in the yield data actually biases downwards our estimates of supply elasticity, suggesting estimates here represent a lower bound. There are undergoing efforts by Deines et al. (2019); Dado et al. (2019) to improve the accuracy of the satellite predictions, and these will be used in the analysis once available. A second limitation of this analysis is that the cross-price elasticities could not be estimated. This was due to a high collinearity between the corn and soybean prices, whose ratio remained relatively stable over the period considered, 2008-2017. Ironically, the years just before and just after this period witnessed much stronger price variations, due to the price spike in 2007, as well as the trade war with China and its impact on soybeans in 2019. While using data per 2008 is difficult due to the unavailability of crop maps for certain states prior to 2008, using data from 2019 will be done in a near future. A final limitation of the current analysis is that spatial heterogeneity in the response could not be investigated. Preliminary results indicate however important heterogeneity, in line with results about area response from Pates and Hendricks (2018). This seems an interesting line of research, and I leave it for future work.

A Derivations

A.1 Derivation of equation 1

Note that notation here differs slightly from the notation in the main text. Yield function y has to be substituted for f, density and cumulative functions $f()$ and $F()$ become $g()$ and $G()$. Finally, the main text describes the case where land above $\theta^\star$ produces, while the proof below is for the case where fields below $\theta^\star$ do produce.

We want the derivative of $\bar{f}(p) = \int_0^{\theta^\star(p)} g(\theta)f(p,\theta)d\theta/G(\theta^\star(p))$ with respect to prices. Using Leibniz rule for the integral, together with the ratio rule, leads to:

$$\frac{\partial \bar{f}(p)}{\partial p} = \frac{\left[g(\theta^\star(p))\, f(p,\theta^\star(p))\frac{d\theta^\star(p)}{dp} + \int_0^{\theta^\star} g(\theta)\frac{\partial f(p,\theta)}{\partial p}d\theta\right]\cdot G(\theta^\star) + \int_0^{\theta^\star} g(\theta)f(p,\theta)d\theta \cdot g(\theta^\star)\frac{d\theta^\star(p)}{dp}}{G(\theta^\star)^2}$$

The first and third terms in the numerator can be combined into (omitting the dependency of $\theta^\star$ on p):

$$\begin{aligned} g(\theta^\star)\frac{d\theta^\star(p)}{dp}\left[f(p,\theta^\star)G(\theta^\star) - \int_0^{\theta^\star} g(\theta)f(p,\theta)d\theta\right] &= \\ g(\theta^\star)\frac{d\theta^\star(p)}{dp}\left[\int_0^{\theta^\star} g(\theta)f(p,\theta^\star)d\theta - \int_0^{\theta^\star} g(\theta)f(p,\theta)d\theta\right] &= \\ g(\theta^\star)\frac{d\theta^\star(p)}{dp}\left[\int_0^{\theta^\star} g(\theta)[f(p,\theta^\star) - f(p,\theta)]d\theta\right] & \end{aligned}$$

Bringing this term back into the main equation leads to:

$$\frac{\partial \bar{f}(p)}{\partial p} = \frac{\int_0^{\theta^\star} g(\theta)\frac{\partial f(p,\theta)}{\partial p}d\theta G(\theta^\star) + g(\theta^\star(p))\frac{d\theta^\star(p)}{dp}\left[\int_0^{\theta^\star} g(\theta)[f(p,\theta^\star) - f(p,\theta)]d\theta\right]}{G(\theta^\star)^2}$$

A.2 Figures

References

Angrist, J. D. and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

BEDDOW, J. M. AND P. G. PARDEY (2015): "Moving Matters: The Effect of Location on Crop Production," *The Journal of Economic History*, 75, 219–249.

BORYAN, C., Z. YANG, R. MUELLER, AND M. CRAIG (2011): "Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program," *Geocarto International*, 26, 341–358.

BURKE, M. AND D. LOBELL (2017): "Satellite-based assessment of yield variation and its determinants in smallholder African systems," *PNAS*, 114:9, 2189–2194.

DADO, W., J. M. DEINES, AND D. B. LOBELL (2019): "Improving satellite-based soybean yield mapping across irrigated and rain-fed conditions," in *American Geophysical Union, Fall Meeting*.

DEINES, J. M., W. DADO, R. PATEL, AND D. B. LOBELL (2019): "Insights into Effective Satellite Crop Yield Estimation from an Extensive Ground Truth Dataset in the US Corn Belt," in *American Geophysical Union Fall Meeting*.

EHRLICH, P. R. AND J. HARTE (2015): "Opinion: To feed the world in 2050 will require a global revolution," *Proceedings of the National Academy of Sciences*, 112, 14743–14744.

EL-NAZER, T. AND B. A. MCCARL (1986): "The Choice of Crop Rotation: A Modeling Approach and Case Study," *American Journal of Agricultural Economics*, 68, 127–136.

FARMAHA, B. S., K. M. ESKRIDGE, H. Y. K. G. CASSMAN, J. E. SPECHT, AND P. GRASSINI (2016a): "Rotation Impact on On-Farm Yield and Input-Use Efficiency in High-Yield Irrigated Maize-Soybean Systems," *Agronomy Journal*, 108:23, 2313–2321.

FARMAHA, B. S., D. B. LOBELL, K. E. BOONE, K. G. CASSMAN, H. S. YANG, AND P. GRASSINI (2016b): "Contribution of persistent factors to yield gaps in high-yield irrigated maize," *Field Crops Research*, 186, 124 – 132.

GARDNER, B. (1976): "Futures Prices in Supply Analysis," *American Journal of Agricultural Economics*, 58, 81–84.

GODFRAY, H. C. J., J. R. BEDDINGTON, I. R. CRUTE, L. HADDAD, D. LAWRENCE, J. F. MUIR, J. PRETTY, S. ROBINSON, S. M. THOMAS, AND C. TOULMIN (2010): "Food Security: The Challenge of Feeding 9 Billion People," *Science*, 327, 812–818.

HAILE, M. G., M. KALKUHL, AND J. VON BRAUN (2014): "Inter- and intra-seasonal crop acreage response to international food prices and implications of volatility," *Agricultural Economics*, 45, 693–710.

——— (2016): "Worldwide Acreage and Yield Response to International Price Change and Volatility: A Dynamic Panel Data Analysis for Wheat, Rice, Corn, and Soybeans," *American Journal of Agricultural Economics*, 98, 172–190.

HENDRICKS, N. P., J. P. JANZEN, AND A. SMITH (2015): "Futures Prices in Supply Analysis: Are Instrumental Variables Necessary?" *American Journal of Agricultural Economics*, 97, 22–39.

HENDRICKS, N. P., S. SINNATHAMBY, K. DOUGLAS-MANKIN, A. SMITH, D. A. SUMNER, AND D. H. EARN-

HART (2014a): "The environmental effects of crop price increases: Nitrogen losses in the U.S. Corn Belt," *Journal of Environmental Economics and Management*, 68, 507 – 526.

HENDRICKS, N. P., A. SMITH, AND D. A. SUMNER (2014b): "Crop Supply Dynamics and the Illusion of Partial Adjustment," *American Journal of Agricultural Economics*, 96, 1469–1491.

HENNESSY, D. A. (2006): "On Monoculture and the Structure of Crop Rotations," *American Journal of Agricultural Economics*, 88, 900.

JIN, Z., G. AZZARI, AND D. B. LOBELL (2017): "Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches," *Agricultural and Forest Meteorology*, 247, 207 – 220.

JOHNSTON, C. A. (2014): "Agricultural expansion: land use shell game in the U.S. Northern Plains," *Landscape Ecology*, 29, 81–95.

LARK, T. J., J. M. SALMON, AND H. K. GIBBS (2015): "Cropland expansion outpaces agricultural and biofuel policies in the United States," *Environmental Research Letters*, 10, 044003.

LIVINGSTON, M., M. J. ROBERTS, AND Y. ZHANG (2015): "Optimal Sequential Plantings of Corn and Soybeans Under Price Uncertainty," *American Journal of Agricultural Economics*, 97, 855.

LOBELL, D. B. AND G. AZZARI (2017): "Satellite detection of rising maize yield heterogeneity in the U.S. Midwest," *Environmental Research Letters*, 12, 014014.

LOBELL, D. B., D. THAU, C. SEIFERT, E. ENGLE, AND B. LITTLE (2015): "A scalable satellite-based crop yield mapper," *Remote Sensing of Environment*, 164, 324 – 333.

LYBBERT, T. J., A. SMITH, AND D. A. SUMNER (2014): "Weather Shocks And Inter-Hemispheric Supply Responses: Implications For Climate Change Effects On Global Food Markets," *Climate Change Economics (CCE)*, 5, 1450010–1–1.

MACEWAN, D. AND R. HOWITT (2011): "An Empirical Model of Crop Rotations," Tech. rep., Prepared Paper for the 55 th Annual AARES Conference.

MAGRINI, E., J. BALIÉ, AND C. MORALES-OPAZO (2018): "Price Signals and Supply Responses for Staple Food Crops in Sub-Saharan Africa," *Applied Economic Perspectives and Policy*, 40, 276–296.

MCNEW, K. AND D. GRIFFITH (2005): "Measuring the Impact of Ethanol Plants on Local Grain Prices," *Review of Agricultural Economics*, 27, 164–180.

MIAO, R., M. KHANNA, AND H. HUANG (2016): "Responsiveness of Crop Yield and Acreage to Prices and Climate," *American Journal of Agricultural Economics*, 98, 191–211.

MOTAMED, M., L. MCPHAIL, AND R. WILLIAMS (2016): "Corn Area Response to Local Ethanol Markets in the United States: A Grid Cell Level Analysis," *American Journal of Agricultural Economics*, 98, 726.

MUSSER, W. N., V. J. ALEXANDER, B. V. TEW, AND D. A. SMITTLE (1985): "A Mathematical Programming Model For Vegetable Rotations," *Southern Journal of Agricultural Economics*, 17.

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

PATES, N. J. AND N. P. HENDRICKS (2018): "Estimating Heterogeneous Corn Price Response in the United States," Tech. rep., Department of Agricultural Economics, Kansas State University.

PLOURDE, J. D., B. C. PIJANOWSKI, AND B. K. PEKIN (2013): "Evidence for increased monoculture cropping in the Central United States," *Agriculture, Ecosystems & Environment*, 165, 50 – 59.

PORTER, P., J. LAUER, W. LUESCHEN, J. FORD, T. HOVERSTAD, E. OPLINGER, AND R. CROOKSTON (1997): "Environment affects the corn and soybean rotation effect," *Agronomy Journal*, 89, 441–448.

REN, J., J. B. CAMPBELL, AND Y. SHAO (2016): "Spatial and temporal dimensions of agricultural land use changes, 2001-2012, East-Central Iowa," *Agricultural Systems*, 148, 149 – 158.

RICARDO, D. (1821): *On the Principles of Political Economy and Taxation*, Library of Economics and Liberty.

ROBERTS, M. J. AND W. SCHLENKER (2013): "Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate," *American Economic Review*, 103, 2265–2295.

ROBERTS, M. J., W. SCHLENKER, AND J. EYER (2013): "Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change," *American Journal of Agricultural Economics*, 95, 236.

SCHLENKER, W. AND M. J. ROBERTS (2006): "Nonlinear Effects of Weather on Corn Yields*," *Applied Economic Perspectives and Policy*, 28, 391.

——— (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.

SEARCHINGER, T., R. HEIMLICH, R. A. HOUGHTON, F. DONG, A. ELOBEID, J. FABIOSA, S. TOKGOZ, D. HAYES, AND T.-H. YU (2008): "Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change," *Science*, 319, 1238–1240.

STERN, A., P. DORAISWAMY, AND R. HUNT (2012): "Changes of crop rotation in Iowa determined from the United States Department of Agriculture, National Agricultural Statistics Service cropland data layer product," *Journal of Applied Remote Sensing*, 6, 1 – 16 – 16.

STEVENS, A. (2015): "Fueling Local Water Pollution: Ethanol Refineries, Land Use, and Nitrate Runoff," Tech. rep., University of Berkeley.

STIGLER, M. (2019a): "Measuring rotation effects in the US Corn Belt," Tech. rep., Chapter 2 of dissertation, `https://github.com/MatthieuStigler/MatthieuStigler.github.io/raw/master/docs/rotation_effects_Stigler_standalone.pdf`.

——— (2019b): "US Corn Belt: a satellite view," Tech. rep., Chapter 1 of PhD dissertation, `https://github.com/MatthieuStigler/MatthieuStigler.github.io/raw/master/docs/Chapter1_Stylised_facts_standalone.pdf`.

THOMAS, A. (2003): "A dynamic model of on-farm integrated nitrogen management," *European Review of Agricultural Economics*, 30, 439.

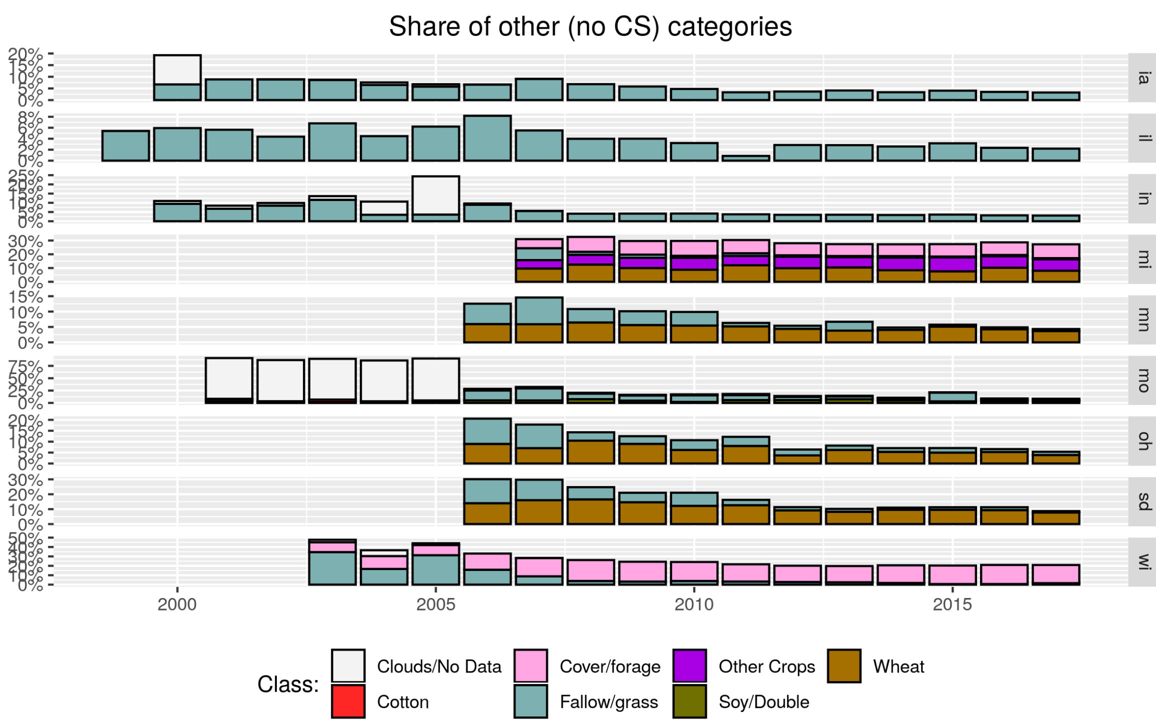
THORNTON, P., M. THORNTON, B. MAYER, Y. WEI, R. DEVARAKONDA, R.S.VOSE, AND R. COOK. (2017): "Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version3." ORNL DAAC, Oak Ridge, Tennessee, USA.

TILMAN, D., C. BALZER, J. HILL, AND B. L. BEFORT (2011): "Global food demand and the sustainable intensification of agriculture," *Proceedings of the National Academy of Sciences*, 108, 20260–20264.

WEERSINK, A., J. H. CABAS, AND E. OLALE (2010): "Acreage Response to Weather, Yield, and Price," *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 58, 57–72.

WRIGHT, C. K. AND M. C. WIMBERLY (2013): "Recent land use change in the Western Corn Belt threatens grasslands and wetlands," *Proceedings of the National Academy of Sciences*, 110, 4134–4139.

Figure 10: Shares of other crops, by state and year



Note: shares based on the fields available retained in the dataset.

Figure 11: Corn and soybeans location

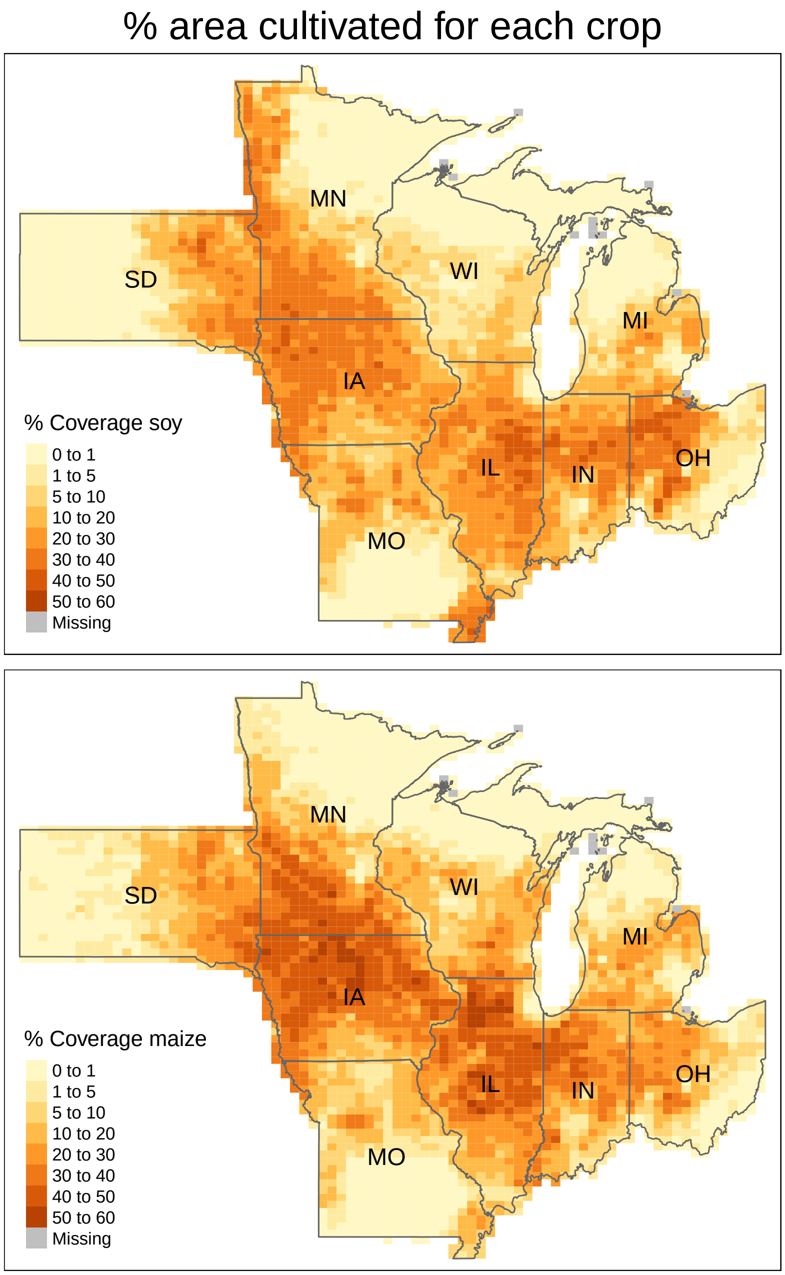


Figure 12: Corn yield response to fertiliser

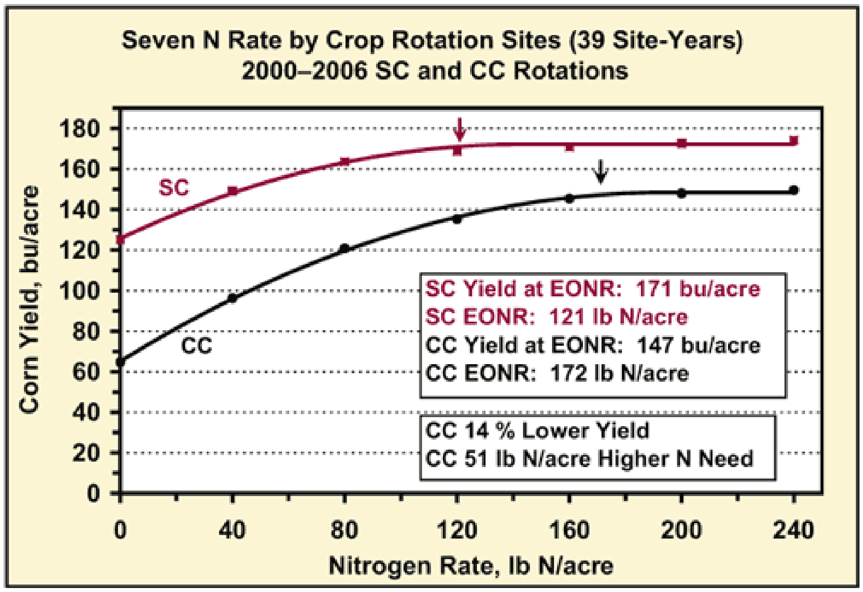


Figure 13: Conditional rotation history, 2000-2010