

Soil Data Not Considered in Cornerstone U.S. Agricultural Policy

Joshua D. Woodard*
Cornell University

Leslie J. Verteramo Chiu
Cornell University

February, 2016

Working Draft



**Author Information and Disclosure: Dr. Joshua D. Woodard is the Zaitz Family Faculty Fellow of Agribusiness and Finance and assistant professor in the Dyson School of Applied Economics and Management at Cornell University. Leslie J. Verteramo Chiu is a postdoctoral research assistant. This work was partially supported by the Meridian Institute. Woodard is the founder of Ag-Analytics.org, is an expert reviewer for the Federal Crop Insurance Corporation Board and United States Department of Agriculture to review plans of insurance, serves on the AGree Conservation and Crop Insurance Task Force (CCIT), and has developed crop insurance rating systems for several products currently sold in the market. Parts of this work were presented in September, 2015 at a joint USDA/ERS-Farm Foundation Soil Health Workshop. Joshua Woodard takes responsibility for all views and opinions expressed. Corresponding Author Email: jdw277@cornell.edu*

Copyright Joshua D. Woodard, 2015. Ag-Analytics.Org

Soil Data Not Considered in Cornerstone U.S. Agricultural Policy

Abstract

As of the Agricultural Act of 2014, the Federal Crop Insurance Program (FCIP) is now the cornerstone agricultural policy in the United States, and is the largest such program globally with about \$100 billion in coverage annually. Given its scale and scope, it has the potential to have pervasive impacts on incentives and policy functioning if not designed and priced properly. Surprisingly, soil data are not considered by the Government when establishing insurance guarantees or rates. Using soil data which could easily and feasibly be scaled nationally, we find that the pricing differentials caused by the Government's failure to handle soil information leads to large errors. Policy implications are discussed.

Introduction

The Federal Crop Insurance Program is now the cornerstone agricultural policy in the United States. Despite the occasional anti-crop insurance opinion piece, the reality is that Federal Crop Insurance and related risk management programs are of utmost importance to extant domestic agricultural policy. The Federal Crop Insurance Program maintains volumes of nearly \$10 billion in premium annually alone in recent years, on around \$100 billion dollars of coverage (Woodard, 2013). While the bulk of program criticism tends to frame the discussion around what the benefits are for farmers and crop insurance companies, the Program is actually priced, regulated, and administered by the Federal Government via the little known "Risk Management Agency" (RMA) of the United States Department of Agriculture (USDA). Private companies deliver the program under terms and premiums rates at the behest of the RMA, and farmers optionally participate.

Surprisingly, the Government does not utilize soil data in designing products, rates, or setting guarantees. Risk management programs such as the Federal Crop Insurance will likely only become more important in coming decades as farmers and food security are faced with increased risk from market volatility and climate change (Battisti, and Naylor, 2009; Lobell et al., 2011; Lobell, Schlenker, and Costa-Roberts, 2011; Lobell, Sibley, and Ortiz-Monasterio, 2012; Lobell et al., 2013; Lobell, et al., 2014; Lesk, Rowhani, and Ramankutty, 2016). These programs also have the potential to have pervasive impacts on conservation outcomes if not designed or priced appropriately, but may enable sustainability objectives if properly structured (Woodard et al., 2012a).

There is a large literature on the modeling of crop yield distributions in the agricultural economics and actuarial literature (see e.g., Woodard, Sherrick, and Schnitkey, 2011; Woodard and Sherrick, 2012; Woodard, 2014), but due to data limitations and/or Government data suppression, most studies tend to lack explicit consideration of soil and site specific data on policy and insurance design. While evaluating the effects of soil on crop growth is, of course, very common in crop sciences on small scales and in trial work, very little has been done on integrating these data and approaches for the purposes of large scale insurance estimation in public policy contexts, with few exceptions (e.g., Woodard, 2014; Woodard, 2015a). Legally, each insurance contract sold should be priced so that the amount indemnified in expectation equals the premium set by the Government. Nevertheless, attaining this type of pricing efficiency via a Government agency, as many have pointed out, is likely an unrealistic task for a government agency, and is likely more suited for market determination (not only in crop insurance, but also generally; see e.g., Priest, 1996; Cummins, 2006; Jaffee, 2006; Michel-Kerjan and Kousky, 2010).

Instead of using soil data to determine baseline insured yield levels and premium rates, the Government's methodology relies on a noisy measure of average historical yields which does not account for the number nor specific years of production reported, the weather in those years across different farmers' policies, nor even which fields being insured. Thus, the Government's method does not reflect full information regarding soils, or for example when a producer adds or removes new land from an insured unit. This can result in mispricing of the underlying insurance and misalignment of incentives.

The purpose of this study is to investigate the feasibility of using high resolution soil data for the modeling of crop insurance guarantees in large scale contexts for this cornerstone agricultural program, and the implications of omitting the same relative to current Government methods (for the first time, to our knowledge). The soil data and soil health indexes employed must be nationwide and have high availability in order to be scalable and operational in practice; thus we employ the SURRGO soil dataset from National Resources Conservation Service (NRCS), which is high resolution, nationwide soil type data (of which there are several thousand soil types). While we acknowledge that the resolutions at which the data are published (10-30 meter) may lend themselves to a mild degree of inaccuracy, this study evaluates at the Common Land Unit (roughly speaking, field boundary level) which typically consists of several acres and has a reasonable scientific basis (NRCS, 2015).¹ Fifty seven different off-the-shelf soil quality attributes such as available water storage, soil organic carbon, and other aggregate soil quality indexes, are matched with the soil type data to map soil types into quantifiable indexes in estimating field level yield guarantee models.

¹ Regardless, it is difficult to argue that in this "big data" era, that soil information should not getting tracked, and updated via the administration of programs like Federal Crop Insurance.

Employing highly available yield and crop cover data, soil conditional expected yield models are estimated and downscaled to the field level in order to estimate the distribution of impacts on rating errors stemming from omitting soil information. For this study, we focus exclusively on how omitting soil information in determining *baseline insurable yields* (i.e., “*guarantees*”) leads to pricing errors. Heterogeneity in soil conditional risk and intra-county premium rate differentials may also, in reality, vary with soil type (see Woodard, 2015a); thus, the true premium error in existing Federal Crop Insurance Program (FCIP) rates from omitting this information is likely larger than that presented here.

Results indicate that pricing efficiency could be significantly improved in this program by taking into account soil data explicitly when estimating crop insurance guarantees. This could lead to savings for taxpayers, fairer premiums for lower risk farmers, less risky underwriting exposure for companies and the taxpayers, and possibly better environmental outcomes. This framework could be applied and expanded in a fairly straightforward manner to other field level yield databases (which, since 2009 the RMA has started to collect, but does not utilize) to operationalize in practice. Before it will be feasible to operationalize the adaptation of the FCIP to more properly accommodate emerging conservation practices, having rating systems in place with soil information at the foundation will be critical and necessary. That said, it should be recognized that the FCIP interacts with other policies (e.g. Commodity Title programs that are insurance-like), markets, not to mention the Standard Reinsurance Agreement. Thus, one might argue that future changes to the rating system to incorporate soil data should be carefully thought out and considered in close concert with stakeholders, delivery companies, and policy-makers (and perhaps not agency staff), as opposed to a unilateral and rushed response by the RMA to respond with something in reaction to new research.

Results

Figure 1 displays corn acres planted by county for 2014, and Figure 2 displays the average National Commodity Crop Productivity Index for corn and soybeans (NCCPI) by county. Soil quality is highly spatially correlated and varies widely across the U.S., from 0.15 (5th percentile of corn producing counties) in marginal counties regions to 0.75 (95th percentile) in intensive production regions in the heart of the Midwest. The standard deviation across counties is about 0.187. Not surprisingly, soil quality tends to match the density of planted acres closely. Figure 3 displays the NCCPI for McLean County, Illinois (the largest county by acreage in the state in one of the most productive areas) by Common Land Unit (CLU) field boundary; within the county, substantial variation in soil quality is still very apparent; the average NCCPI value across all fields is in McLean is 0.73, and the standard deviation is 0.139 (0.443 and 0.896 at the 5th and 95th percentiles).

The crop insurance program does not track individual field location in determining guarantees, but rather uses an average of between 4 and 10 years of producer data (depending on how much is available); this average does not take into account which years are reported, and so two fields that have very similar soil could get much different guarantees simply by virtue of which years are reported; likewise, it is not unusual for lower quality fields to obtain higher guarantees than higher quality fields simply because of which years were reported. How overall quality of soil within a policy changes from adding or removing land within a unit is also not factored into the guarantee.

To evaluate how this intra-county soil variation translates into expected errors in determining guarantees, and resultant impact on rates, we estimate soil conditional expected

yield models using a nation-wide yield data. Regression results for a model using data from the Central Corn Belt states for are displayed in Table 1. See Supplementary Appendix for description of data and additional models to evaluate robustness. The soil parameters are economically and statistically significant. The first model uses *Soil Organic Carbon*, and the second model includes *Root Zone Available Water Storage*. These variables represent soil quality proxies. The third model uses the NCCPI. The measure of *Soil Organic Carbon* was taken for the layer between 20 and 50 cm, and corresponds to units of grams of carbon per square meter (NRCS, 2015). This layer provided the best response to corn yields, although NCCPI reveals similar (stronger) results. Weather controls and state level fixed effects were also taken into account and found to be significant.

Figure 1-Acres Planted, Corn, 2014

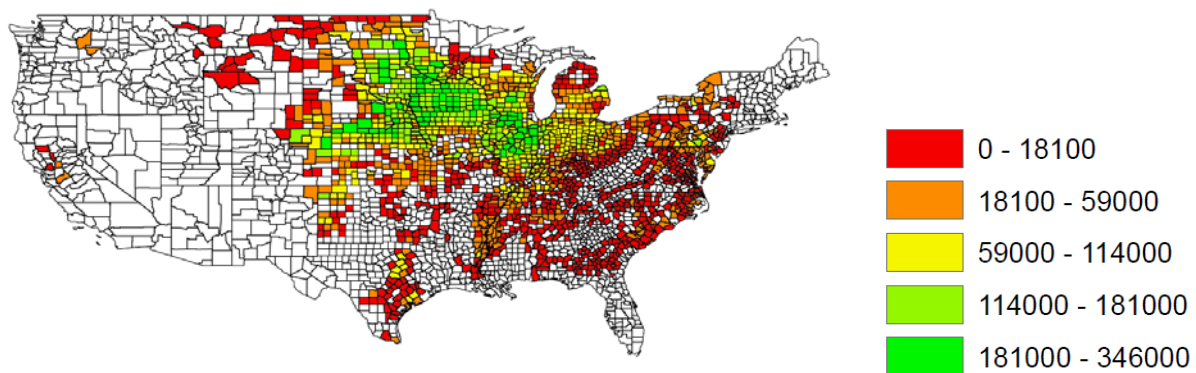


Figure 2-National Commodity Crop Productivity Index, Corn, United States by County

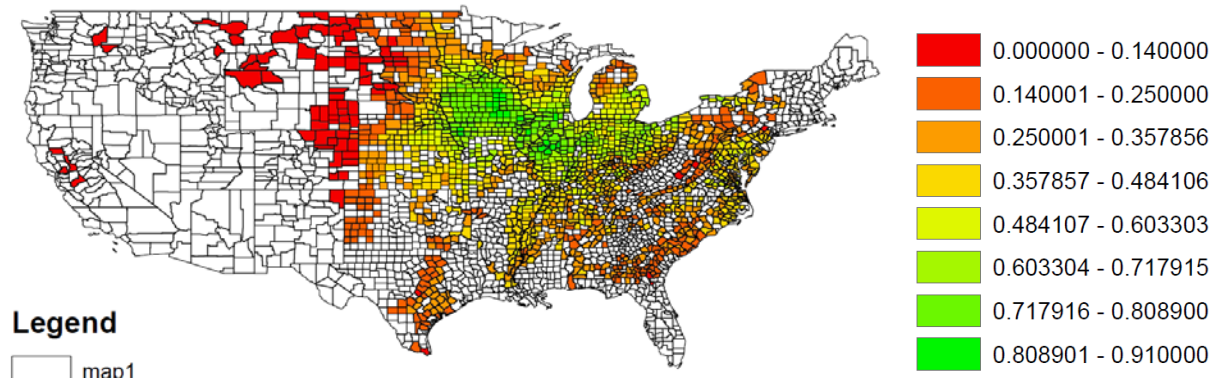


Figure 3-National Commodity Crop Productivity Index by Field, McLean County, Illinois

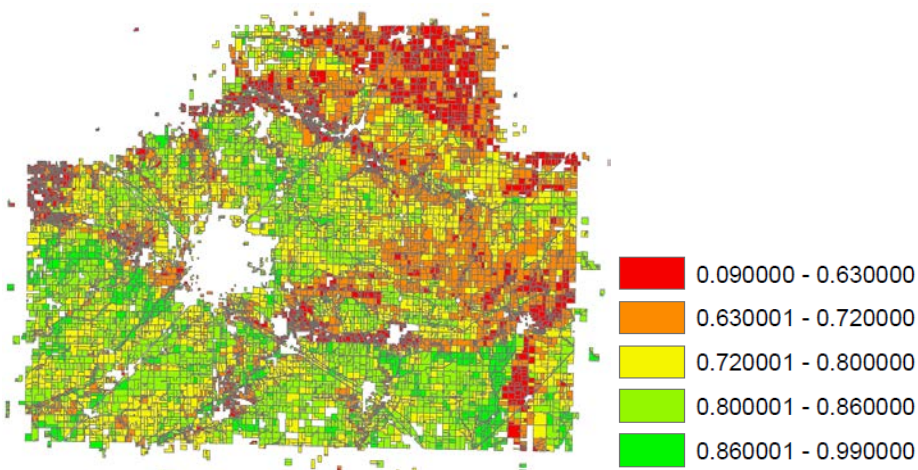


Table 1 – Yield Regression Results, State-Level Fixed Effects, Soil, and Weather Controls

<i>Variable</i>	<i>Model 1</i>		<i>Model 2</i>		<i>Model 3</i>	
	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>
<i>IL</i>	-456.883	-16.26	-436.836	-16.29	-434.083	-16.18
<i>IN</i>	-463.182	-16.47	-439.837	-16.39	-433.746	-16.15
<i>IA</i>	-471.315	-16.77	-455.120	-16.97	-446.69	-16.65
<i>TimeIL</i>	1.815	67.81	1.813	71.02	1.802	70.53
<i>TimeIN</i>	1.626	57.49	1.631	60.45	1.625	60.17
<i>TimeIA</i>	2.054	76.58	2.052	80.23	2.048	80.00
<i>Temp</i>	50.999	21.12	46.629	20.22	44.919	19.43
<i>Prec</i>	0.732	42.09	0.748	45.05	0.747	44.94
<i>Temp2</i>	-1.252	-24.16	-1.163	-23.50	-1.129	-22.75
<i>Prec2</i>	-0.0027	-38.86	-0.0027	-41.65	-0.0027	-41.92
<i>SOC</i>	0.00198	19.92	0.00153	16.02		
<i>Rootz</i>			0.1503	33.50		
<i>NCCPI</i>					80.1144	39.12
Adj.R ²	0.659		0.690		0.689	
SSR	345.66		314.41		315.08	
AIC	66035		64967		64989	
DW	1.46		1.59		1.57	

Simulation experiments were conducted using the estimated yield models--downscaling to the field level, which can be conducted without bias in the *mean* yield--for two different cases (*own-field* and *mixed-field*, below). Refer to the SI for further information on estimating premium rates and rate error. In the first, between 4 and 10 yields were simulated for each field from each field's own estimated yield distribution. The guarantee (so called, APH) and expected loss rate was then calculated by integrating over the yield distribution (see SI). This design allows for evaluation of the likely inefficiency that results from the small sample nature of determining guarantees and ignoring soil information. In the second experiment (to account for the fact that in reality is not only soil taken into account, but the APH is not necessarily pegged to an individual field), we first select at random a field in the county, then randomly draw

between 4 and 10 years of yield data to determine APH; we then impose that APH on another field at random, and calculate the expected loss rate under that field's yield distribution. Each experiment was run for 3,524,500 iterations (500 iterations per field for 7049 fields).

Table 2 presents results of the rate simulation analysis. Figure 5 reports kernel densities of the simulated relative pricing error multiples for experiment 1 (*own-field*) and 2 (*mixed-field*). Relative and nominal rate errors are displayed by percentile for all iterations. Nominal rate error was calculated as quoted rate minus the true expected loss rate once accounting for inefficiency in APH measure from omitting soil; a value of less than 0 indicates underpricing, while a value greater than 0 indicates overpricing. Relative rate error was calculated as the expected loss rate divided by the quoted rate; a value of less than 1 indicates overpricing, while a value greater than 1 indicates underpricing. Standard deviation of *own-field* and *mixed-field* nominal rate error was 0.0095 and 0.0203, respectively. Coefficient of variation of *own-field* and *mixed-field* nominal rate error was 54.7% and 113.71%, respectively. The results show the potential for substantial mispricing to stem from inefficient guarantee determination. In 25 percent of all simulated cases (assuming the APH is not mixed across fields), the expected loss rate to the charged premium rate would be less than 0.745, or the equivalent of an overcharge in premium of approximately 34%). For the lowest risk 10% of cases, they are approximately double-charged. On the other hand, in the riskiest 10% of cases, the policies would pay out 1.851 times the premium charged, meaning that the premium would need to be 185.1% of the current value charged to be actuarially fair.

Table 2- Rate Error Simulation Results (Percentiles) by Field; McLean County Illinois

	<i>Percentile</i>						
	<i>5th</i>	<i>10th</i>	<i>25th</i>	<i>50th</i>	<i>75th</i>	<i>90th</i>	<i>95th</i>
<i>Own-Field, Relative</i>	0.417	0.526	0.745	1.047	1.426	1.851	2.153
<i>Mixed-Field, Relative</i>	0.352	0.528	0.958	1.696	2.927	4.880	6.661
<i>Own-Field, Nominal</i>	-0.020	-0.015	-0.007	-0.001	0.004	0.008	0.010
<i>Cross-Field, Nominal</i>	-0.105	-0.070	-0.034	-0.012	0.001	0.008	0.011

Figure 4 displays the average relative pricing error *by field*. Only fields as indicated by the 2013 Cropland Data Layer maps as having a majority of corn were considered for the analysis (although results were consistent regardless). Note, values less than one indicate that on average the field will be overpriced and vice-a-versa for values greater than one. Comparing to the soil quality map in Figure 3, it is clear that higher quality fields will be over-priced substantially, lower quality fields underpriced. The rate error multiples are also large. For example, the lowest 1/3 of fields by soil quality had estimated rate error multiples of 2.349 or greater, meaning that the expected loss on the field under the RMA's rating methodology would result in indemnities about 2.3 times greater than premium. On the other hand, the highest 10% quality fields would, on average, be overpriced. Note that this analysis is under the conservative assumption that RMA rates are, for a given true expected yield level, in fact correct. To the extent that the conditional loss rate is more responsive across APH levels than indicated by RMA's methodology, then the rate error levels could be even higher.

Figure 4 - Average Pricing Error Multiple (*Mixed Field*), by Field; McLean County, Illinois

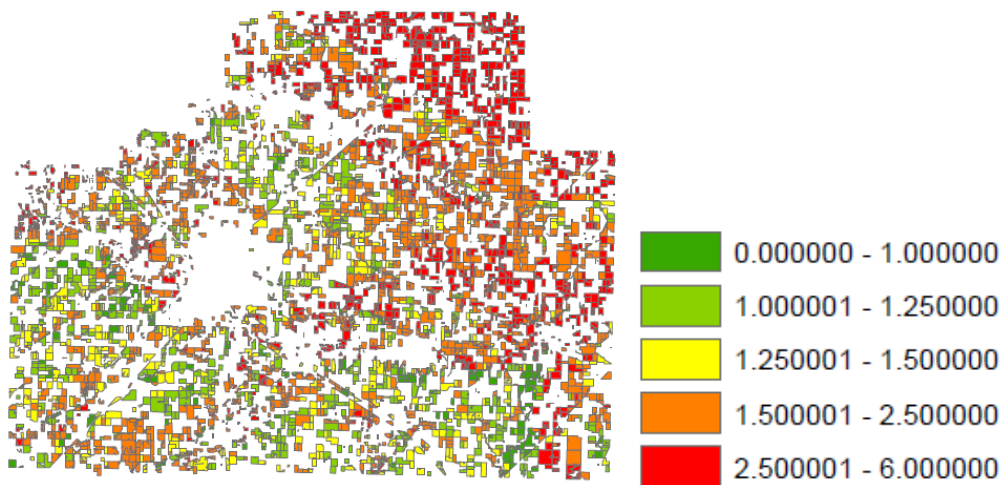
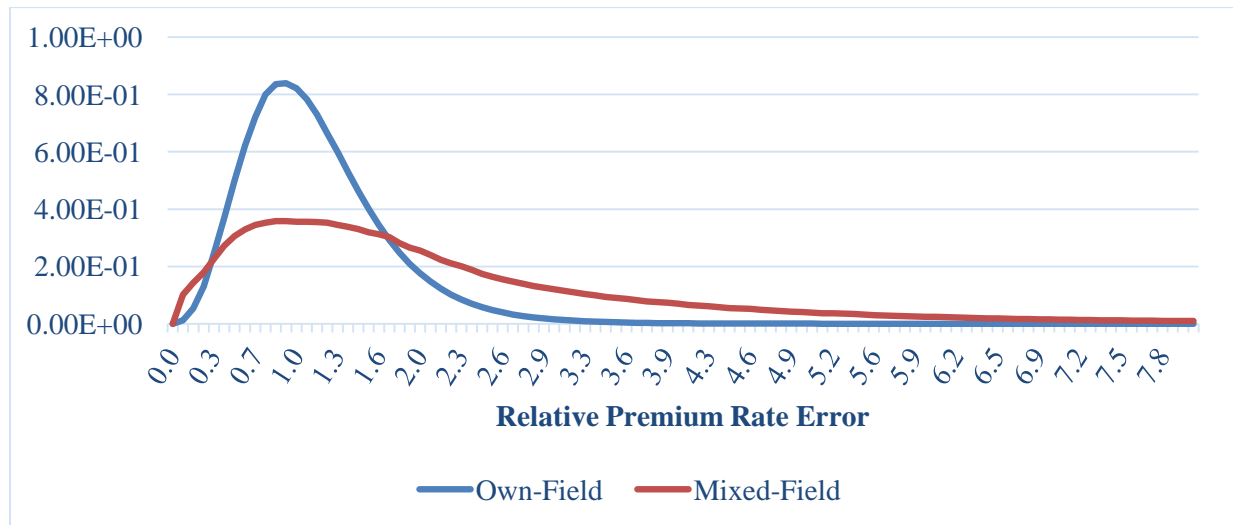


Figure 5-Kernel Density, Simulated Relative Pricing Error Multiples; McLean, Illinois



Discussion and Conclusion

We have shown that widely available high resolution soil data can feasibly be integrated into crop insurance guarantee determination, and that a large boost in rating efficiency is likely relative to currently employed rating methods used by the Government which ignore explicit field location and soil data. While the data exist, RMA does not use this information in estimating premium rates. This study serves as a proof of concept and provides a strong motivation for modifying the program to incorporate these data.

The availability of insurance affects investment and production decisions. Past research has shown that poorly designed insurance can also lead to adverse incentives regarding practice choice (Woodard et al., 2012a). Approaches to integrating soil information explicitly into yield risk and insurance models is a necessary precursor to later accommodating and quantify impacts from different soil sustainability practices in dynamic frameworks. For example, the amount of soil organic carbon (SOC) can be modified through managerial practices. These practices include the adoption of no-till farming, cover crops, and reforestation. They are commonly referred to as sustainable agricultural practices as they increase SOC, while also increasing soil fertility/productivity and sequestering carbon. This increase of *SOC*, however, is not immediate, but contributes to the long-term productivity of soils. Increased soil organic matter may also contribute to water retention in the long run (Arriaga, 2003, pg. 889). That said, in certain conditions they can lead to higher risk, and thus should be considered carefully.

For example, some practices may have complex impacts that also vary by soil type or climatology. Increasing the use of cover crops can potentially result in higher soil water retention in some soils (Scott et al., 1990; Teasdale and Mohler, 1993). However, in arid and semi-arid

regions, cover crops may have negative effects on soil water retention since they compete with the commodity crop for available water (Snapp et al., 2005; Unger and Vigil, 1998). Usually these types of factors are not adequately accounted for in the pricing and administration of the program by the RMA.

Building a soils-based pricing foundation is a first step towards creating a crop insurance system which can accommodate future program modifications related to sustainability. This would open the door for improving conservation outcomes by appropriately incentivizing (or at least not disincentivizing) adoption via insurance which is appropriately designed and rated. Failure to properly determine guarantees under this program may lead producers to not adopt otherwise potentially optimal conservation practices such as cover crop use, skip-row (Woodard et al., 2012a), adaptive nitrogen management (van Es et al., 2007), or others which they might contrarily. Nevertheless, the current structure of how insurance rules are determined, and how insurance is priced (i.e., via an Agency, as opposed to market discovery of rates), leaves the system unable to respond flexibly.

A lack of understanding surrounding the shortcomings of the current Actual Production History (APH) methods employed--as well as program rigidities--may hamper adoption of these newer and more appropriate approaches. Since 2009, the Government began collecting insurance policy records and yields tied to CLU's, but to date simply does not make much use of the data, nor is it clear if they have managed it appropriately.² In order to scale up the analysis to the entire country for the purposes of *operationalizing modifications to the program* to take into account soil quality explicitly, the Government would likely need to release proper data to allow for further research and development by companies and the research community. To date, RMA

² Specifically, the Common Land Units (CLU's, or roughly speaking, fields) within each Insured Unit.

has refused to make these data available to the research community, citing privacy concerns. We urge the government to make these data available to researchers in a suitable form. Future data-intensive research applications will take on an increasingly larger role in shaping policy and sustainability science going forward (Woodard, 2015b). The fact that these data exist, but that viable researchers may not access them, provides a strong motivation for the creation of a secure data warehousing facility to foster data-intensive analytics which otherwise cannot be conducted without the large troves of uncured data whose access is restricted (sometimes reasonably, and sometimes not) by various government agencies. These important program modifications could have overarching conservation and sustainability impacts.

There is fairly broad agreement and much empirical evidence (in crop insurance and other fields) that government agencies are likely not in a natural position to be jointly administering, regulating, and pricing these types of insurance programs, and indeed the track record in pricing efficiency is not favorable. Agencies tend to be reactive rather than proactive, and often lack sufficient bandwidth to suitably replicate pricing mechanisms that markets would yield (see e.g., Priest, 1996; Cummins, 2006; Jaffee, 2006; Michel-Kerjan and Kousky, 2010; Woodard et al., 2012b). This makes a very strong case for replacing the current government controlled rating system with a more realistic, accurate, and flexible set of pricing mechanisms where companies and the market can have some hand in informing premium rates. This would allow the market to flexibly integrate information and data which a government agency cannot adequately handle, but which have a foundational importance in the classification of the risks to be underwritten (e.g., the influence of different soil types on crop growth). The Federal Crop Insurance Program as a policy has made many inroads into helping farmers manage risk, and it is well accepted that such markets would likely not exist in the absence of government intervention

due to systemic risk. Much like deposit insurance, terrorism insurance or even health insurance, government intervention and subsidization likely has a legitimate role to play in economically optimal outcomes; however, the Government arguably should not be in charge of actually setting prices/rates for the program.

References

- Arriaga, F. J., Lowery, B., “Soil Physical Properties and Crop Productivity of an Eroded Soil Amended with Cattle Manure” *Soil Science*, Volume 168(12), December 2003, pp 888-899
- Battisti, D. S. & Naylor, R. L. Historical warnings of future food insecurity with unprecedented seasonal heat. *Science* 323, 240–244 (2009)
- Coumou, D. & Rahmstorf, S. A decade of weather extremes. *Nature Clim. Change* 2, 491–496 (2012)
- Cummins, J. D., 2006, Should the Government Provide Insurance for Catastrophes? Federal Reserve Bank of St. Louis Review, 88(4): 337-379.
- Jaffee, D., 2006, Should the Government Provide Insurance for Catastrophes? Comments on a Paper by J. David Cummins, Federal Reserve Bank of St. Louis Review, 88(4): 337-379.
- Krueger, E. S., Ochsner, T. E., Porter, P. M., & Baker, J. M. (2011). Winter rye cover crop management influences on soil water, soil nitrate, and corn development. *Agronomy Journal*, 103(2), 316-323.
- Lesk, C., Pedram Rowhani, and Navin Ramankutty, “Influence of extreme weather disasters on global crop production,” *Nature* 529, 84–87 (07 January 2016).
- Lobell, D. B. et al. The critical role of extreme heat for maize production in the United States. *Nature Climate Change* 3, 497–501 (2013)
- Lobell, D. B., Sibley, A. & Ivan Ortiz-Monasterio, J. Extreme heat effects on wheat senescence in India. *Nature Climate Change* 2, 186–189 (2012)
- Lobell, D. B., Bänziger, M., Magorokosho, C. & Vivek, B. “Nonlinear heat effects on African maize as evidenced by historical yield trials,” *Nature Climate Change* 1, 42–45 (2011)
- Lobell, D. B., Schlenker, W. & Costa-Roberts, J., “Climate trends and global crop production since 1980.” *Science* 333, 616–620 (2011)
- Lobell, D. B. et al. “Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest,” *Science* 344, 516–519 (2014)

McIsaac, G.F., David, M.B., Gertner, G.Z., and Goolsby, D.A. "Relating net nitrogen input in the Mississippi River basin to nitrate flux in the lower Mississippi River: A comparison of approaches." *Journal of Environmental Quality*, 31, 1610, 2002.

Michel-Kerjan, E. O., and C. Kousky, 2010, Come Rain or Shine: Evidence on Flood Insurance Purchases in Florida, *Journal of Risk and Insurance*, 77(2): 369-397.

Miranda, M. J., and J. W. Glauber, 1997, Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets, *American Journal of Agricultural Economics*, 79(1): 206-215.

Natural Resources Conservation Service, United States Department of Agriculture, Soil Survey Staff, Web Soil Survey. Available online at <http://websoilsurvey.nrcs.usda.gov/>. Accessed May, 2014.

Priest, G. L., 1996, The Government, the Market, and the Problem of Catastrophic Loss, *Journal of Risk and Uncertainty*, 12: 219-237.

Snapp, S. S., S. M. Swinton, R. Labarta, D. Mutch, J. R. Black, R. Leep, J. Nyiraneza, and K. O'Neil. "Evaluating cover crops for benefits, costs and performance within cropping system niches." *Agronomy Journal* 97, no. 1 (2005): 322-332.

Teasdale, J. R., Mohler, C. L. "Light Transmittance, Soil Temperature, and Soil Moisture under Residue of Hairy Vetch and Rye" *Agronomy Journal*, Vol. 85 no. 3, 1993.

Unger, Paul W., and Merle F. Vigil. "Cover crop effects on soil water relationships." *Journal of Soil and Water Conservation* 53, no. 3 (1998): 200-207.

van Es, H.M., Czymmek, K.J., and Ketterings, Q.M. "Management effects on N leaching and guidelines for an N leaching index in New York," *Journal of Soil Water Conservation* 57, 499, 2002.

van Es, H.M., Kay, B.D., Melkonian, J.J., and Sogbedji, J.M. Nitrogen management under maize in humid regions: Case for a dynamic approach. In: Bruulsema, T. (ed.). *Managing Crop Nutrition for Weather*. International Plant Nutrition Institute Publ., Norcross, GA, 2007, pp. 6–13.

Woodard, J., Sherrick, B. J., & Schnitkey, G. D. (2011). Actuarial Impacts of Loss Cost Ratio Ratemaking in U.S. Crop Insurance Programs. *Journal of Agricultural and Resource Economics*. 36:211-228.

Woodard, J., & Sherrick, B. J. (2011). Estimation of Mixture Models using Cross-Validation Optimization: Implications for Crop Yield Distribution Modeling. *American Journal of Agricultural Economics*. 93:968-982.

Woodard, J., Pavlista, A. D., Schnitkey, G. D., Ward, K., & Burgener, P. (2012a). Government Insurance Program Design, Incentive Effects, and Technology Adoption: The Case of Skip-Row Crop Insurance. *American Journal of Agricultural Economics*. 94:823-837.

Woodard, J., Schnitkey, G. D., Sherrick, B. J., Lozano-Gracia, N., & Anselin, L. (2012b). A Spatial Econometric Analysis of Loss Experience in the U.S. Crop Insurance Program. *Journal of Risk and Insurance*. 79:261-186.

Woodard, J.D., "Theme Overview: Current Issues in Risk Management and U.S. Agricultural Policy," *Choices* (Sept. 2013). (Article page views as of 12/2013: 491; Theme Page views: 2609)

Woodard, J. (2014). A Conditional Distribution Approach to Assessing the Impact of Weather Sample Heterogeneity on Yield Risk Estimation and Crop Insurance Ratemaking. *North American Actuarial Journal*.

Woodard, J.D., "Integrating High Resolution Soil Data into Federal Crop Insurance Rates: Actuarial, Policy, and Sustainability Implications," *Working Paper* 2015a.

Woodard, J.D., "Data Management, Large Scale Empirical Applications, and "Big Data" in Agricultural Economics, Policy, and Climate Change," *Ag-Analytics.Org Concept Document*, Cornell University, 2015b.

Selected Supplemental Information

Data and Methods

Yield data were collected from the National Agricultural Statistical Service (NASS) for corn, and span 39 years from 1975 to 2013 for 8 states: IL, IN, IA, MI, MN, MO, OH, and WI. These states were selected because they are the largest corn producing states by total production, form a contiguous area, and have similar crop growing seasons ($C=734$ counties, $N=26,755$ observations).

Growing season weather data were obtained from the PRISM Climate Group at Oregon State University. Average monthly temperature and precipitation were obtained from the PRISM Climate Group at Oregon State University. PRISM data are gridded 4km resolution data, which we aggregate to obtain county averages. Weather values for July and August are employed as explanatory variables, as these months are critical for corn production in this region. These serve as reasonable proxies for this analysis.

This study employs the last published Common Land Unit (CLU) field maps in the public domain (last released in 2008 by the Farm Service Agency). While the exact field boundaries change somewhat through time as units are combined, sold, added, or broken out, we use them here to approximate the degree of spatial heterogeneity in soils across fields within a county. The analysis could easily be replicated with updated CLU maps. Since the 2008 Farm Bill, under an unfunded mandate, RMA began directing participating companies (so called Approved Insurance Providers, or AIP's) to collect CLU specific yield and insurance policy data. Prior to that, data in the crop insurance program were only identified by the county in which the policy was located. Thus, the analysis contained herein should be fairly extensible to internal RMA field level data.

Soil data were obtained from the National Resources Conservation Services (NRCS) SSURGO database. NRCS publishes a gridded version of this dataset at 30 meter resolution, along with the VALU1 table to link soil type to 57 different soil quality indexes. These soil variables are generally stated for different soil layers (depths) for different soil components, including soil organic carbon (*SOC*), average water storage (*AWS*), and thickness of soil components. There are also indexes for the National Commodity Crop Productivity Index (*NCCPI*) for various crops, Drought vulnerable soil landscapes (*DROUGHTY*), and Potential Wetland Soil Landscapes (*PWSL*), and root zone depth available water storage (*rootz*). We spatially aggregate over the areas of interest to obtain 57 average soil quality estimates at both the county and CLU levels of aggregation for modeling. To ensure that the soil averages for the counties were constructed only for soils upon which corn is typically grown, we filtered the SSURGO data by the NASS Cropland Data Layers (CDL), averaging only over areas grown to corn. The CDL layers provide 30 meter resolution estimates of crop cover type. All processed data and sources are freely available from the Ag-Analytics open-data platform (Woodard, 2015; online at ag-analytics.org).

SOC is an important determinant in soil fertility. *rootz* is the volume of plant available water that the soil can store within the root zone and is also an important determinant in yield potential. Since most variables within each group captures the same factors, they are expected to be highly correlated and may pose problems if included in the regression. We selected the variables that best explain crop productivity by running successive regressions of yields on each soil variable, and evaluating the size of the average effect. We selected two variables for further analysis below with the largest average effects from each class: *SOC*, measured in g C/m^2 in standard layer 3 (20 to 50 cm depth), and *rootz*, expressed in mm. These variables are both

positively correlated with crop yields, and have a correlation of about 0.34 between them. We also generate results below using the *NCCPI* (corn and soybeans)—which is an aggregate proxy for soil quality—for comparison. Note that at the 10m resolution, the SSURGO soil data may be subject to some false precision, but the CLU's are typically several acres, which is less of a concern. Clearly though, having a system in place to update soil data through the administration of the FCIP would be advantageous, and but would be a later step.

Tables S1-S3 below report summary statistics for the sample yield (S1), weather (S2) and soil data employed (S3). Several regression models of yield on soil and weather are estimated. Models are estimated with state-level fixed effects for both intercept and trend terms. Models were also estimated using a subset of the data for Illinois, Indiana, and Iowa only for robustness. Several variants of the models are investigated, including those with and without weather/soil interactions, as well as with different subsets of soil variables for robustness. Explicitly, Tables S4-S6 display supplementary regression results to evaluate robustness of the yield models, including models that incorporate soil and weather interactions for the three Central Corn Belt states of Illinois, Indiana, and Iowa (S4), models for the larger 8 state region (S5, state fixed effects, including soil and weather controls), and the 8 state region also including weather and soil interactions (S6). The results across all models were fairly consistent, and use of any of the below models would not affect the main rate analysis discussed in the main body of the text in context.

Insurance Analysis and Simulation Design

Using the estimated yield model which employs the NCCPI, field level expected yields for McLean County were estimated. These can be downscaled to the field level without bias since

they measure the first moment. Actual RMA rates were collected for each field based on the expected yield to perform a counter-factual analysis. The RMA rating methodology assigns a premium rate within a county depending on the APH level (which is a proxy of expected yield). The mean RMA quoted rate (which we take as true for the purposes of the counter-factual analysis) was 0.0173 for this analysis, with a standard deviation of 0.00089. Using those rates, Weibull distribution parameters were backed out for each field by estimating the parameter values which would result in the premium rate quoted by RMA. In general, the premium rate is set equal to the expected loss cost ratio, explicitly,

$$Premium\ Rate = \frac{\int_0^{E(Y) \cdot Cov} Max(0, APH \cdot Cov - y) \cdot f(y) dy}{(APH \cdot Cov)} \quad \textbf{Equation (1)}$$

where Cov is the coverage level election (or one minus the deductible percent), $f(y)$ is the yield distribution, y is the yield outcome, and APH is expected yield; in this context, the product of $APH \cdot Cov$, is the insurance guarantee. Simply put, the actuarially fair premium rate, or expected loss cost, is the expectation of indemnities (or losses) weighted over all possible yield outcomes as a fraction of the amount of coverage (or guarantee). This analysis focuses on results for the 85% coverage level (the highest election available), although the same contextual results hold for any coverage level (and in fact the relative error is greater for lower coverage levels due to the convexity of the expected loss curve with respect to the coverage level).

Table S1. 2009-2013 Average Corn Yields (bu./acre)

State	Mean	Std	Min	Max
IL	147.74	37.23	19.00	199.80
IN	145.04	34.99	30.10	211.50
IA	157.65	29.22	44.50	207.00
MI	135.66	24.63	48.50	181.60
MN	157.41	24.94	75.00	195.40
MO	114.22	36.88	22.30	192.40
OH	153.95	23.82	66.10	195.00
WI	142.16	22.17	60.70	182.00
Sample	144.74	33.35	19.00	211.50

Table S2. Average Temperature and Precipitation (1975-2013)

State	Temperature				Precipitation			
	Mean	Std	Min	Max	Mean	Std	Min	Max
IL	23.90	1.57	18.72	28.51	93.99	33.79	21.82	238.88
IN	23.17	1.41	19.53	27.56	101.77	33.77	27.59	245.18
IA	22.75	1.44	17.77	27.51	104.78	45.26	11.21	320.12
MI	20.48	1.53	15.30	24.41	84.31	23.42	33.00	191.77
MN	20.98	1.47	15.58	24.90	94.76	33.58	17.51	258.55
MO	25.04	1.37	20.94	29.81	97.83	39.41	16.74	306.91
OH	22.39	1.18	19.17	26.13	99.18	30.55	35.75	231.74
WI	20.63	1.50	15.21	24.49	101.25	32.16	32.13	263.95
Sample	22.62	2.08	15.21	29.81	97.58	35.58	11.21	320.12

Table S3. Summary Statistics for SOC and AWS. SOC is measured in g C/m² in standard layer 3 (20 to 50 cm depth), AWS is measured in mm in standard zone 5 (0 to 150 cm depth).

State	SOC				AWS			
	Mean	Std	Min	Max	Mean	Std	Min	Max
IL	5,904.86	2,008.12	2,297.81	9,779.49	239.73	23.20	181.55	280.23
IN	6,388.47	2,266.47	3,260.61	15,339.95	211.52	29.76	126.48	277.71
IA	8,547.06	2,304.71	4,639.72	13,279.93	276.59	14.61	239.20	310.11
MI	5,786.77	1,718.51	2,788.87	19,660.04	196.57	33.30	78.59	257.13
MN	8,669.02	2,602.61	2,584.34	13,359.41	223.99	43.30	98.86	287.55
MO	4,601.17	1,192.74	2,153.28	7,786.98	232.11	30.24	126.44	274.47
OH	5,279.71	1,507.02	2,911.91	9,098.12	191.22	21.95	131.10	255.85
WI	5,314.41	1,640.88	2,852.49	10,930.51	201.65	37.07	129.80	287.91
Sample	6,336.40	2,431.10	2,153.30	19,660.00	224.50	39.95	78.59	310.11

Table S4. Yield Regressions - Eight States with Soil and Weather Variables

<i>Variable</i>	<i>Model 4</i>		<i>Model 5</i>		<i>Model 6</i>	
	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>
<i>IL</i>	-507.210	-44.78	-483.779	-43.88	-417.133	-37.49
<i>IN</i>	-512.031	-45.09	-486.047	-43.96	-416.209	-37.24
<i>IA</i>	-523.208	-46.21	-502.048	-45.57	-428.146	-38.45
<i>MI</i>	-534.004	-47.66	-507.528	-46.51	-433.566	-39.22
<i>MN</i>	-546.089	-48.53	-523.205	-47.77	-440.817	-39.68
<i>MO</i>	-519.566	-46.14	-494.889	-45.14	-421.61	-38.00
<i>OH</i>	-517.207	-45.47	-489.934	-44.23	-425.578	-38.06
<i>WI</i>	-527.246	-46.89	-501.742	-45.83	-428.036	-38.61
<i>TimeIL</i>	1.822	64.59	1.818	66.29	1.800	66.06
<i>TimeIN</i>	1.630	54.55	1.632	56.20	1.624	56.26
<i>TimeIA</i>	2.083	73.66	2.077	75.56	2.061	75.46
<i>TimeMI</i>	1.699	49.69	1.676	50.44	1.625	49.19
<i>TimeMN</i>	2.384	75.20	2.384	77.38	2.353	76.83
<i>TimeMO</i>	1.621	56.83	1.597	57.57	1.630	59.16
<i>TimeOH</i>	1.616	52.60	1.625	54.43	1.631	54.99
<i>TimeWI</i>	1.732	49.68	1.724	50.90	1.695	50.35
<i>Temp</i>	52.418	51.79	48.670	49.25	42.229	42.06
<i>Prec</i>	0.7672	58.26	0.7707	60.22	0.7467	58.70
<i>Temp2</i>	-1.2365	-54.95	-1.1652	-53.10	-1.0395	-46.84
<i>Prec2</i>	-0.00273	-49.96	-0.00277	-52.02	-0.00273	-51.62
<i>SOC</i>	0.00324	44.65	0.00248	33.99		
<i>Rootz</i>			0.12315	39.51		
<i>NCCPI</i>					74.35916	63.70
Adj. R2	0.6536		0.6727		0.6768	
SSR	388.47		367.05		362.44	
AIC	159560		158050		157710	
DW	1.35		1.39		1.36	

N= 26755

Table S5. Yield Regressions - Eight States with Soil and Weather Variables Interactions

<i>Variable</i>	<i>Model 7</i>		<i>Model 8</i>	
	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>
<i>IL</i>	-524.8002	-44.46	-457.663	-41.07
<i>IN</i>	-527.5552	-44.58	-458.311	-40.91
<i>IA</i>	-542.0921	-46.01	-468.923	-42.05
<i>MI</i>	-547.8125	-46.95	-473.706	-42.79
<i>MN</i>	-564.3944	-48.05	-482.413	-43.33
<i>MO</i>	-536.5075	-45.61	-466.259	-41.86
<i>OH</i>	-531.3276	-44.83	-468.267	-41.77
<i>WI</i>	-542.5202	-46.32	-469.348	-42.22
<i>TimeIL</i>	1.813	66.37	1.7913	66.40
<i>TimeIN</i>	1.641	56.68	1.6184	56.75
<i>TimeIA</i>	2.0455	74.43	2.0347	75.18
<i>TimeMI</i>	1.6838	50.77	1.5679	47.93
<i>TimeMN</i>	2.3631	76.92	2.3312	77.03
<i>TimeMO</i>	1.6033	58.01	1.6349	60.07
<i>TimeOH</i>	1.6355	54.96	1.6393	55.93
<i>TimeWI</i>	1.7391	51.48	1.6693	50.06
<i>Temp</i>	50.4064	49.95	39.198	39.25
<i>Prec</i>	0.9192	47.55	0.773	41.54
<i>Temp2</i>	-1.1914	-51.88	-0.8181	-34.76
<i>Prec2</i>	-0.0026	-48.74	-0.0026	-48.80
<i>Temp*SOC</i>	-0.0004	-10.67		
<i>Prec*SOC</i>	0.00000	-3.92		
<i>Temp*Rootz</i>	0.0035	2.52		
<i>Prec*Rootz</i>	-0.0007	-8.53		
<i>SOC</i>	0.0119	13.72		
<i>Rootz</i>	0.1103	3.31		
<i>Temp* NCCPI</i>			-11.7458	-25.8865
<i>Prec* NCCPI</i>			-0.0771	-3.0224
<i>NCCPI</i>			338.1252	32.1762
Adj. R2	0.6756		0.6847	
SSR	363.82		353.60	
AIC	157820		157050	
DW	1.37		1.3754	

N= 26755

Table 4. Yield Regressions-Three States Fixed Effects with Soil and Weather Interactions.

	<i>Model 9</i>		<i>Model 10</i>		<i>Model 11</i>	
<i>Variable</i>	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>	<i>Coefficient</i>	<i>T-value</i>
<i>IL</i>	-500.4923	-13.64	-442.8347	-11.51	21.759	10.33
<i>IN</i>	-503.9415	-13.75	-442.9726	-11.52	31.091	16.31
<i>IA</i>	-518.2541	-14.12	-455.2866	-11.83	15.025	6.78
<i>TimeIL</i>	1.8208	71.78	1.7979	70.47	1.621	48.00
<i>TimeIN</i>	1.6572	61.60	1.6405	60.69	1.500	41.87
<i>TimeIA</i>	2.0272	79.50	2.0394	79.70	1.984	58.57
<i>Temp</i>	47.9908	16.78	43.5799	16.06		
<i>Prec</i>	1.0487	35.42	0.9909	26.11		
<i>Temp2</i>	-1.1617	-19.88	-1.1008	-21.57		
<i>Prec2</i>	-0.0026	-39.39	-0.0027	-40.40		
<i>Temp*Soc</i>	-0.0001	-1.37				
<i>Prec*Soc</i>	0.0000	-4.25				
<i>Temp*rootz</i>	-0.0044	-1.52				
<i>Prec*rootz</i>	-0.0012	-10.38				
<i>SOC</i>	0.0044	2.94				
<i>Rootz</i>	0.3704	5.13				
<i>Temp*NCCPI</i>			0.0533	0.04		
<i>Prec*NCCPI</i>			-0.3533	-7.11		
<i>NCCPI</i>			113.5153	3.75	102.144	37.87
Adj.R ²	0.6947		0.6909		0.4469	
SSR	309.79		313.63		561.19	
AIC	64808		64941		71499	
DW	1.58		1.57		1.76	

N= 11293