Integrating High Resolution Soil Data into Federal Crop Insurance Policy: Implications for Policy and Conservation

Joshua D. Woodard

Author Affiliation: Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, 14853 USA

Email: jdw277@cornell.edu

Author Vitae:

Joshua D. Woodard is an assistant professor and the Zaitz Family Faculty Fellow of Agribusiness and Finance in the Dyson School of Applied Economics and Management at Cornell University. 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 (CCITF), and has developed crop insurance rating systems for several products currently sold in the market.

Abbreviations: Risk Management Agency (RMA), Federal Crop Insurance Corporation (FCIC), Actual Production History (APH), Common Land Units (CLUs), Farm Service Agency (FSA), Soil Productivity Rating (SPR)

Abstract

Well-designed and executed policies are critical for aligning sustainability incentives and enabling future agricultural productivity growth. In the U.S., government-administered crop insurance is the primary direct mechanism through which agriculture is subsidized and represents over \$100 billion in liabilities annually. Despite the importance of soil properties in determining crop yield formation and risk, the Government does not consider any soil information in generating premium rates under the Federal Crop Insurance Program. The purpose of this study is to investigate the potential of integrating high-resolution soil data into modeling of field-level insurance rates in large-scale applications. Here, using the actual distribution of soil quality across crop fields in a high production region, models are developed to incorporate soil data into insurance rates and then evaluated to investigate the magnitude of risk differentials across different soil qualities. These soil-conditioned results were then compared to rates that would have been generated by the Government's current soil-naïve methodology. This study indicates that the degree to which soils vary within a county is highly significant, leading to rating errors of 200% or greater. Implications of ignoring soil information and operational considerations of modifying this cornerstone program are discussed.

1. Introduction

Maintenance of soil health is critical to ensuring the future of agricultural productivity and adapting to climate change. Likewise, well-designed policies are necessary for fostering appropriate production incentives and accommodating innovations in conservation and sustainability. The "Big Data" revolution has led to an increased interest in exploring opportunities to employ high-resolution data in large scale policy applications to improve sustainability of the agricultural system, which previously were impractical or impossible. Given the pervasiveness of economic policy on production decisions, it is important to have a sound understanding of how agricultural policy in major markets impacts incentives.

In the U.S., government-administered crop insurance is the primary direct mechanism through which agriculture is subsidized and represents over \$100 billion in liabilities annually (Woodard, 2013). The Federal Crop Insurance Program (FCIP) is a public-private partnership. The Federal Government--via the United States Department of Agriculture Risk Management Agency (RMA) and Federal Crop Insurance Corporation (FCIC)--is solely responsible for determining the set of insurance products, rules, subsidies, and rates/prices for the insurance. The program works by insuring farmers against yield or revenue losses in the event of low prices, drought, or other perils. Private companies sell the insurance and bear the primary financial risk,

but do not have authority over rate-setting or rule-making procedures. Insurers must sell insurance to any producer who wishes to buy it, under the prices and rules set by the government, and in return the Government provides a partial layer of reinsurance coverage to the insurance companies. Farmers pay part of the premium, and participation is optional. Farmers also have the option of insuring at different coverage levels/deductibles (with higher coverages having higher premium rates). Premium rates vary based on region, crop, coverage level, and product type.

Of fundamental importance to any insurance market is the ability to set insurance rates that properly reflect risk, and by law the Government should attempt to price premiums as actuarially fair so that the expected insurance loss payments equate to the unsubsidized premium (Woodard et al., 2012a). Failure to do so can lead to what is known as adverse selection, whereby the insured may be incentivized to only buy insurance when they have a high likelihood of collecting a claim, due to possessing some risk factor not being observed by the insurer or rate setting authority (in this case, the Federal Government). Lower risk producers will then be less likely to participate, which can lead to higher rates over time, higher and less predictable insurer losses, low participation, higher taxpayer costs, and less efficient markets. Relatedly, inefficient rating of risk can lead to adverse incentives. For example, severely mispriced crop insurance could potentially result in producers choosing to produce on higher risk or environmentally sensitive land they might not otherwise (Lubowski et al., 2006). There have been documented cases in which poorly designed government insurance policy can lead to adverse incentives regarding which management practices producers adopt, potentially dis-incentivizing conservation-oriented cropping practices (Woodard et al., 2012b).

Soil type is an important factor in determining crop yield potential and subsequently would be expected to affect insurance losses. Previous work has highlighted the importance of such intra-regional yield variability (e.g., Claassen and Just, 2011; Lobell, Ortiz-Monasterio, and Falcon, 2007; Popp, Rudstrom, and Manning, 2005; Woodard, 2014), yet little work has been done on evaluating the impact of soil on probabilistic yield models in large-scale, high resolution contexts. Surprisingly, soil information is not utilized by the Government when determining premium rates under the current U.S. crop insurance program, and no study has yet linked high-resolution soil data back to field-level yield data explicitly on large scales for evaluating or pricing insurance. The RMA's rating procedures rely on a noisy measure of average historical yields across mixed fields as the primary basis for classifying rates within a county. Thus, rates do not reflect full information regarding soils, specific years reported, nor even fields reported. In 2009, the RMA began collecting data matched to individual contiguous farming parcels, Common Land Units (CLUs, or roughly speaking, fields), and reached 100% target reporting by 2016. High-resolution soil type data are also readily available in the U.S.; however, despite availability of these data, the information is yet un-used by RMA for pricing this program.

The purpose of this study is to investigate the potential of integrating high-resolution soil data into the modeling of micro-level (field) yield distributions and insurance rate-setting for

large-scale applications. Here, using the actual distribution of soil quality across crop fields in a high production region, models are developed to incorporate soil data into insurance rates. Results are evaluated to assess the magnitude of risk differentials across different soil qualities and compared to rates that would have been generated by the Government's current soil-naïve methodology. Implications for policy and operational considerations are also discussed.

2. Methods and Data

High resolution soil type data from the SURRGO soil dataset (Natural Resources Conservation Service, 2015), along with Common Land Unit (CLU) field boundary maps maintained by the Farm Service Agency (FSA), are matched with yield models estimated from a large multi-decade field level dataset consisting of over 120,000 annual field level observations covering 1980-2008 to conduct an insurance rating analysis for the state of Illinois for corn. Statistical crop yield models are estimated to obtain soil and location conditional yield distributions and insurance rates. These models are then overlaid onto the CLU field maps maintained by the Farm Service Agency (FSA), and compared to the rates which the current RMA rating system would generate in order to evaluate the magnitude and importance of omitting soil information in administering the crop insurance program. The documentation regarding the RMA rating system is contained in various data files and publications on RMAs website; an installation of this system was constructed by the authors, and a public web API is available online at *Ag-Analytics.Org* (Woodard, 2016a, Woodard, 2016b). The premise of the methodology is to first construct estimates of what RMA would have charged for insurance on different fields/farms, and then to estimate yield models conditional on soil data to estimate actual risk of loss (i.e., insurance premium rates) for those fields; these implied loss rates which take into account soil are then compared to the RMAs published rates. Cropland Data Layers from the National Agricultural Statistics Service are also employed to filter land units for the specified crop to ensure evaluation is conducted only on fields where corn crops are grown. Next an explanation of the insurance structure is provided, followed by a discussion of the estimation of crop yield distributions and their use in rating (or pricing) the insurance.

2.1 Crop Insurance Product Structure

Crops insurance is structured such that it pays an indemnity when realized yields or revenue is below some elected coverage level multiplied by a yield or revenue baseline. The coverage level is equal to one minus the deductible percent, and typically ranges between 50%-85%, with higher coverage levels demanding a higher premium rate; the farmer elects the coverage level amount, and pays more for insurance with a higher coverage level. The amount the farmer pays for coverage to the insurer is known as the premium. Payments for losses to the farmer are called indemnities. The baseline yield against which losses are evaluated, known as the Actual Production History (*APH*), is an average of between the last 4 and 10 years of data (depending on how many years of reportable data the producer possesses). Explicitly, the standard yield

insurance contract works by insuring yield losses, with the payment per acre of land insured (indemnity/acre) calculated as:

$$\text{Indemnity} / \text{Acre} = \text{Max}(0, \text{APH} \cdot \text{Cov} - y) * \text{BP}$$

where, *Cov* is percent coverage level insured as elected by the producer (between 50%-85%, similar to a deductible), *y* is the end of season yield per-acre (for corn, measured in bushels per acre), and *BP* is the base crop price per unit of the crop being insured (currently around \$4/bushel for corn). Thus, the contract pays the farmer for any eventual yield losses below some known baseline amount (the historical average times the coverage level elected, minus the actual yield at the end of the season) at the stated base price. Note that higher coverage levels imply larger insurance payments when there is a loss, so higher coverage level products also have a higher premiums. For example, if the APH is 100 bu./acre, the BP is \$4.00/bu., and coverage level is 50%, and the eventual yield is 40 bu./acre, then the indemnity would be equal to \$40/acre.

2.2 Yield Distribution Modeling

Statistical yield models are employed for the basis of this study, as is standard in insurance contexts. While mechanistic crop models are useful in many branches of science, direct estimation of yield models using historical data has been argued on empirical grounds in many contexts (see e.g., Peng et al., 2004). Indeed, in insurance applications, empirical/statistical approaches are virtually exclusively employed, and are thus adopted here. Farm level yield data are obtained from the Illinois Farm Business Farm Management dataset from 1980-2008. The data report annual corn yields and acreage by farm along with which county the farm exists. The data also report a field specific soil productivity rating (SPR) using the Illinois Bulletin 810 Soil Series (Olson et al., 2000). The main dataset consisted of 121,416 observations. First, I estimate regression models of crop yield conditional on soil productivity, a time variable in each county to proxy for technology gains, and county level fixed effects (results in Table 1). Use of farm-level fixed effects was also explored, but when including either an SPR or farm-to-county SPR rating, the model was singular, indicating that SPR rating was highly collinear with farm fixed effects, but not necessarily county fixed effects. Next, I evaluate if risk is conditional on soil by restricting the data set to farms with at least 20 years of data, then estimating models of yield standard deviation on soil and acreage, with county fixed effects (results in Table 2). The yield models are employed to generate expected yield and risk measures for each field conditional on the SPR and location. For tractability, I conducted the analysis for McLean County, which is the largest county by corn acreage in Central Illinois. Next, Weibull distributions were constructed for with the desired mean and standard deviation for each CLU using a moment matching approach. The Weibull distribution has consistently been shown to be a valid representation of

corn yields in this region (Woodard and Sherrick, 2011; Woodard, 2014; Woodard, Sherrick and Schnitkey, 2011).

Note that weather as well as overall climatology influences yield outcomes, expected yield, and risk. A number of studies have evaluated these weather impacts in-depth (see e.g., Woodard, 2014; Woodard, Sherrick and Schnitkey, 2011). Here, I proxy for mean differences in expected yield risk using county fixed effects, with the rest of specific weather risk events being relegated to the error term. This avoids the need to perform integration over the joint weather/error/soil distribution to obtain premium rate estimates and does not affect the subsequent soil conditional insurance pricing analysis. This is valid provided soil type is uncorrelated with specific weather outcomes, which must be true in this context since the soil type data are time invariant. Readers are referred to Woodard (2014) for an exposition of such models using these data.

2.3 Soil and Field Boundary Data Preparation

The Illinois Bulletin 810 Soil Series (Olson, et al., 2000) assigns a soil quality rating to each soil type, using a nomenclature which is identical to the NRCS SURGGO soil series. The gridded SURGGO data are nationwide, 10 meter resolution maps of soil type. The goal is to obtain a map of the CLU field boundaries with an average SPR which can be linked back to the yield models. Using the SPR classifications in the Illinois Bulletin 810 (Olson, et al., 2000) which link SURRGO soil type with an SPR rating, a spatial join is conducted between the SURRGO soil layers and the CLU boundary files to obtain an average SPR for each field based on the soils in that field. Last, to ensure only fields where corn is grown are evaluated, I use the Cropland Data Layers (National Agricultural Statistics Service, 2015; Woodard, 2016c) to filter out fields which contain a majority of estimated corn grid cells for analysis. The Cropland Data Layer (CDL) files are 30m resolution map layers published by the USDA which estimate crop cover type. Crop year 2014 CDLs are employed.

2.4 Estimating Field Level Insurance Premium Rates

The job of an actuary is to estimate the expected value of losses arising from such a contingent contract. This is usually expressed by the expectation of the *loss cost ratio* (LCR), which is the ratio of loss (or indemnity paid) divided by the liability (or total value of the commodity being insured). The expected value of the LCR is known as the actuarially fair premium rate. Thus, I estimate soil-conditioned rates of the form,

$$E(LCR | \theta) = \frac{\int_0^{E(Y|\theta) \cdot Cov} \text{Max}(0, E(Y | \theta) \cdot Cov - y) \cdot f(y | \theta) dy}{E(Y | \theta) \cdot Cov}$$

where $f(y | \theta)$ is the Weibull yield density estimated conditional on soil, time, and location (i.e., θ), and $E(Y | \theta)$ is the expected yield conditional on θ which is employed to approximate APH. This provides a reference measure of the underlying true premium rate conditional on soil. The

premium charged per acre is then calculated as liability (value insured) multiplied by premium rate. For comparison, I then generate premium rates from the published online RMA rating system. The RMA rating system takes into account which county the farm is located within, but does not condition on soil. Rather, it uses APH to benchmark premiums within the county. I approximate RMA rates for each field using the expected yield measure generated for each field. Note that this provides conservative estimates of the divergence between the RMA methodology and an explicit soil rating methodology because in reality the APH measure is an inefficient measure of expected yield.

3. Results

3.1 Models of expected yield

Results for the expected yield models are reported in Table 1. The soil productivity rating variable (*SOIL*) was highly significant, as were technology proxies (*TIME*). Results were fairly consistent over four alternative specifications. Models 3 and 4 also include quadratic terms for *SOIL* in order to test for non-linear effects of soil quality on expected yield, but are not significant in this region. Models 2 and 4 also include a quadratic term on *TIME* to evaluate potential non-linear effects of technology gains through time on yields; in both models the quadratic term is negative, indicating positive but concave technology trend curve over the period, though this was insignificant in Model 4, and the effect on this term in Model 2 was not substantial enough to affect the qualitative results of the insurance rating analysis. The adjusted R^2 was around 0.4, with the rest of the variation in yields being attributed to weather or other factors, which is on par with the literature (see e.g., Woodard, Sherrick, and Schnitkey, 2011; Woodard, 2014).

Table 2 reports results for the yield risk models (*Standard Deviation*) which model farm level yield risk as a function of soil quality (*SOIL*) and farm size (*AREA*). In Model 1, Table 2, the *SOIL* coefficient is negative and significant (-0.117), indicating that higher-quality soils in a county were found to have statistically lower risk. This was consistent across all Models. Models 2 and 3 include quadratic terms on soil, and both find that there are significant non-linear effects of soil quality on yield risk. As expected, due to aggregation effects, larger farms also had lower yield variability, as indicated by the *AREA* term. These results are consistent with the results of Woodard (2014) which likewise finds evidence that higher soil quality is related to lower yield risk.

Figure 1 displays estimated expected yields as implied from the models, for each CLU in McLean County using Model 1. By national standards, McLean County is typically regarded to have relatively high soil homogeneity. Even so, large differences are observed in expected yield by soil quality, ranging from 126 bu./acre to 188 bu./acre in expectation. Figure 2 displays a kernel density of the soil rating variable by farm to provide an additional view of how much soil

varies within this region. Similar to the heat map of expected yields, the soils can vary significantly, and there is particularly large tail for poor soils.

3.2 Utilization of soil information in rating insurance

Figure 3 displays results for premium rates for both the soil conditioned rate estimates, and actual RMA rates for 85% coverage yield insurance product (15% deductible). The red line shows the premium rate which RMA's current rating system would charge (for different APH levels), and ranges from about 2.4% to 1.8%. The blue line displays soil conditioned rates which reflect soil information, and reveal expected premium rates actually vary from about 1.5% to 5%. These 2 lines should proxy each other if all information in soils are embedded in rates appropriately. Yet, there are fairly stark differences as illustrated by the results below, suggesting that RMA's rating system is not sufficiently partitioning the classification of risks by soil type (within a county).

3.3 Soil-conditioned versus soil-naïve insurance rates

Finally, tests are conducted to test whether the difference between ratings generated by the RMA methodology (soil-naïve) versus the soil-explicit (soil-conditioned) methodology is statistically significant. For this, I estimate regressions of the soil-explicit rate on the RMA-generated rate. If RMA rates sufficiently capture the information embedded in soil information through the current method of baselining against the APH average yield measure, then the regression coefficient will be insignificantly different from 1.0. Table 3 presents results of these tests for both the 85% coverage level and 65% coverage level. In both cases, resultant parameters are statistically different from 1.0 (values of 4.8 and 2.6, respectively, for the 85% and 65% coverage levels). The slope of the estimated coefficients (both greater than 1.0.) indicate that the RMA rating methodology is not nearly as responsive to soil as it should be if it took into account soil data in estimating actuarially fair rates. These results support the hypothesis that APH measures utilized by the RMA rating methodology insufficiently captures the risk variability present in soils, and their subsequent use in differentiating risk.

The level of risk stratification that exists across soils, once translated into premium rates, is much higher in reality than that implied by the RMA pricing system. Results here indicate that the true premium per acre would be about double what is charged by RMA for low yielding (i.e. higher risk) units in the county of interest. This is likewise true for high yielding (i.e. lower risk) units. Even when making a variety of conservative assumptions that err strongly towards not finding a difference between the two methods, the results clearly support the importance of incorporating soil information into insurance rating methodology in order to properly reflect risk.

4. Discussion

The results indicate that the degree to which soils vary within a county (even for counties considered to have high soil homogeneity in the Central Corn Belt) is highly significant. Implications of ignoring such information in the crop insurance program can be far reaching. Rate differentials in excess of those currently taken into account by the Government's rating system to the tune of more than 200% appear typical even in very conservative scenarios. New approaches that integrate soil information explicitly into yield risk and insurance models are needed to improve the performance and design of this cornerstone agricultural program. These methodology changes are critical to support recent focus on how crop insurance, as a policy vehicle, can better coordinate with environmental goals, accommodate sustainability practices, and promote soil health.

4.1 Implications for Policy, Conservation, Sustainability, and Environmental Outcomes

Historically, the program has been a lightning rod for criticism regarding premium-rate setting procedures, claimed environmental impacts, and within the context of Federal budget debates (see e.g., Government Accountability Office, 2007). Despite such criticisms, the program has also arguably been very successful in terms of inducing increased participation in risk management activities, reducing farm bankruptcies and related transition costs, and reducing legislative moral hazard associated with *ad hoc* disaster assistance (in essence, the agricultural equivalent of "bailouts"). Crop insurance markets tend not to arise in the absence of government intervention, not because they do not serve a useful economic function, but rather because the presence and degree of systemic risk typically precludes the ability of insurers to underwrite without government support (much like, say, Federal Deposit Insurance). Indeed, the case for the systemic risk argument is well supported in agricultural insurance contexts (see e.g., Miranda and Glauber, 1997; Glauber, 2004).

In many ways it could be argued that having a crop insurance program, from a political economy perspective, improves overall welfare versus doing nothing, given that in essence it is a classic case of a missing market. That fact does not absolve the Federal government and related agencies from sound delivery, however, and there is a strong case to be made that soil data should be explicitly taken into account in setting rates. On the other hand, some critics have contended that crop insurance leads to expansion of production on marginal lands which otherwise would likely not be cultivated in the absence of underpriced insurance. To the extent that such claims are true, I would argue that it does not necessarily indicate a shortcoming in crop insurance policy itself, but rather is simply a rating issue which can be ameliorated by modernizing the actuarial systems.

When well-designed, crop insurance can serve to complete missing markets and potentially lead to better alignment of incentives in the agricultural, environmental, market and policy nexus. For example, some research suggests that agricultural producers may in fact over-utilize nitrogen-based fertilizers as a physical form of production "insurance", leading to reduced energy efficiency and unfavorable environmental outcomes; (McIsaac et al., 2002; van Es,

Czymmek, and Ketterings, 2002; van Es et al., 2007). In such cases, appropriately functioning financial insurance markets could provide more favorable environmental outcomes if producers shift from overuse of nitrogen as a form of insurance to financial insurance contracts. Likewise, appropriately designed insurance could lead to alignment of incentives and producer adoption of certain conservation practices in cases where they otherwise might not if access to insurance did not exist, and/or if available insurance contracts did not properly account for such practices (Woodard et al., 2012b).

4.2 Actuarial and Market Efficiency Considerations

There is some evidence that on the extensive margin, if marginal lands do come into production as a result of having mispriced insurance, they would likely be those most prone to environmental sensitivity (Lubowski et al., 2006). Given that the stated goal of the crop insurance program has consistently been articulated as that of maximizing participation, then properly rating risks is a very important policy lever to control in reaching that goal. In this case, if such an important risk factor as soil is not being reflected in rates, then the insurance pool will start to become concentrated in higher risk producers. From year to year, the RMA re-estimates rates based on historical losses. Over time, this will push up average rates, and make the program less attractive to lower risk producers. In order to induce participation by these producers, higher subsidies are then necessary. While subsidies can be argued on a variety of economic grounds and probably could not be eliminated completely in a fully functional program--from an economic-actuarial perspective--on the margin more accurate rates may be a partial substitute for subsidies in generating participation, and as well as more reliably profitable participation.

4.3 Operational and Program Delivery Considerations

Since 2009, the government and participating companies have started collecting field (specifically, Common Land Unit, or CLU) specific yield data in the Federal Crop Insurance Program. Prior to that, data in the crop insurance program and in government surveys were actually not collected and stored at the field level, but rather only identified the county in which the policy was located. This linking of policy lines to CLU's was implemented as an unfunded mandate, although it is unclear to what extent RMA has actively managed those data, or even if consistent standards have been employed and enforced by RMA on such data collection procedures across different companies or in different crops/regions/products.

There are a few ways through which modifications to the crop insurance program can occur in the U.S. The first is an act of Congress. The second is through submission and review by the FCIC (either initiated by RMA or private parties). The RMA also has some administrative authority within boundaries. The procedure which permits members of the public (or RMA) to propose design modifications to the Federal Crop Insurance Program is known as the 508(h) process. While the RMA now collects the information needed to facilitate such modifications,

they have failed to make it available to researchers for this purpose even though this could easily be facilitated under standard arrangements.

A few additional items would need to be addressed to scale-out a soil-linked rating system nationwide, most of which would be fairly minor. The first issue is that the insurance policy data as currently collected are reportedly linked by insured “units” (which could be a collection of known fields) to CLUs. In concept, this simply would mean that the basis of the models would entail aggregating acreage of a handful of fields, which is a trivial extension from aggregating acreage for a single field. It is important to note that the proposal here is to base rates and guarantees on soil type/quality, and region (as well as experience data) but not one-to-one matching through time of an individual field with its loss experience only. Likewise, not all of the historical data that exists will have matching CLU data, but at this point for many of the major regions/crops, more than enough data should exist to explore pilot implementations.

Another operational consideration is that the nationwide SURRGO data would not use the same soil quality index as that employed for this investigation for Illinois (the Bulletin 810; Olson et al., 2000), although the approach remains essentially the same if other soil quality indexes were to be used. There are several off-the-shelf measures published alongside the SURRGO data in the VALU1 dataset (57 in fact, including the National Commodity Crop Productivity Index, Available Water Storage, Soil Organic Carbon, and many others; see Natural Resources Conservation Service, 2015) which could be used when estimating such actuarial models. These data could be used to extend this research to investigate secondary considerations such as distributional form, and higher-order dimensions of soil quality interactions with weather and other factors. The missing link to operationalization, however, is that RMA and FSA make available to researchers the extensive internal yield and updated field boundary data.

The exact manner in which such corrections would be integrated into the RMA’s current rating system would need to be determined, and could possibly be implemented via estimating and conditioning certain existing parameters by soil quality and county. Modifying the current system would likely cause selling companies to bear significant costs in retooling IT systems for delivery and capture of soil quality by field(s), and for training. It is somewhat unknown how well such corrections might perform, however, within the myriad of other problems with the current rating system. The other option would be to finally retire the existing rating system, and redesign from the “ground up”. Modifying the current rating system would be cheaper short-run, but may not be the optimal solution long-run. The existing rating system is quite old, with the foundational aspects stretching back over 30 years. Through time as problems have arisen, the primary approach taken appears to have been to wrap yet another set of corrections around the existing methodology, instead of redesigning the infrastructure. Anticipating how such corrections through time interact when stacked upon each other is typically difficult to control or even track, although a variety of cases have been highlighted in the literature in which logical inconsistencies have arisen in the rating system (e.g., Barnaby, 2007; Woodard et al., 2012a).

Given the foundational nature of soil in evaluating yield risk (and given that in many ways the base rating system is arguably severely outdated), a case could be made that the time is near to replace the existing rating system with a completely new design, with integration of soil information at its foundation. More accurate rates would likely benefit farmers, taxpayers, as well as the participating selling companies, while also assuaging environmental concerns. The current method of determining baseline yields/liabilities (or APH) should also be addressed to reflect soil type as well. This is another large source of potential efficiency gains in addition to those outlined here regarding rates. Similar models could be employed directly to revise the guarantee determination method to obtain more efficient estimates of expected yield.

5. Concluding thoughts and future directions

This report finds direct evidence of the predictive capacity of large-scale, highly available soil data in predicting crop insurance loss rates. This information is not currently reflected in the Government's rating methodology. These rating inefficiencies likely lead to unnecessary adverse selection, thus rendering the program more expensive than it otherwise would be, which in turn exposes the program and makes it more risky for insurers. This matter could be resolved by incorporating soil quality information directly into the insurance rating procedures to increase pricing efficiency.

There may be an inclination to interpret the results here about the Government's failure to integrate this information into the program currently as an indication of policy failure, and then conclude that this is reason to simply dissolve the Federal Crop Insurance Program; I would caution jumping haphazardly to such assertions. While society ought not to absolve the Government from their duty to properly and optimally deliver this program, and foster innovations surrounding the same, advancements in this realm are a normal evolution that are just recently enabled by improved data collection and analytics. Given that the data which will ultimately be necessary to revise the entire Crop Insurance program fall under some privacy and dissemination restrictions--and also reside in several different USDA agencies--a strong case could be made that a secure data research facility should be set up in order to foster this type of collaboration between a data-hungry research and analytics community, and a data-rich set of agencies. More urgent action by the agencies is needed.

As Burney, Davis, and Lobell (2010) point out, improved agricultural productivity is not mutually exclusive with improved sustainability; precision agriculture and improved practices also have a large role to play in both intensification and conservation. I would argue additionally that the design and execution of agricultural financial policy, via the Federal Crop Insurance Program, has an equally profound role in mediating this conservation/intensification dynamic, and that at the foundations of this role are the actuarial systems employed for what is now the cornerstone agricultural program in the United States. There have been questions in the past regarding certain restrictive insurance rules as they apply to conservation practices, such as cover crop use, which arguably have positive soil health impacts. Before one can begin to think about

incorporating corrections into the rating methodology to allow for adoption of some of these conservation practices which currently are not properly accounted for under the Government's methodologies, one should have a system which properly models soil impacts as a prerequisite. A modernized rating infrastructure that can enable integration of soil information into insurance rates is an essential first step towards aligning this major policy vehicle with and enabling sustainability goals.

Acknowledgements

The author is grateful to participants at the Economic Research Service-USDA/Farm Foundation Soil Health Conference 2015, as well as the AGree Initiative and members of the Conservation and Crop Insurance Task Force for helpful comments. The author would also like to thank Diane Wang at Cornell University for her assistance in proofreading and editing the revised manuscript in an effort to appeal to a broader audience. This work was partially supported by the Meridian Institute, and I am greatly indebted for their support.

References

- Barnaby, G. A. "Would You Pay More in Premiums than the Maximum Indemnity Payment?" *Kansas State Research and Extension, Dept. of Agr. Econ., Kansas State University, Manhattan, Kansas* (Sept. 30, 2007). Available online at http://www.agmanager.info/crops/insurance/risk_mgt/rm_html07/ABrates.asp.
- Burney, Jennifer A., Steven J. Davis, and David B. Lobell, "Greenhouse gas mitigation by agricultural intensification," *PNAS* June 29 (2010) 107-26:12052-12057.
- Claasen, R., and R.E. Just, "Heterogeneity and Distributional Form of Farm-Level Yields" *American Journal of Agricultural Economics* (Jan 2011)93-1: 144–160.
- Glauber, J. W., "Crop Insurance Reconsidered," *American Journal of Agricultural Economics*, (2004)86-5: 1179-1195.
- Government Accountability Office (GAO), "Continuing Efforts Are Needed to Improve Program Integrity and Ensure Program Costs Are Reasonable," *Testimony Before the Committee on Oversight and Government Reform, House of Representatives, Statement of Lisa Shames, Acting Director Natural Resources and Environment* (May 3,
- Lobell, David B., Ortiz-Monasterio, I., and Walter P. Falcon, "Yield uncertainty at the field scale evaluated with multi-year satellite data," *Agricultural Systems* (2007)92:76-90.
- Lubowski, R. N., S. Bucholtz, R. Claassen, M. J. Roberts, J. C. Cooper, A. Gueorguieva, and R. Johansson, "Environmental Effects of Agricultural Land-Use Change: The Role of Economics and Policy" *Economic Research Service Report No. 25*. (2006).
- 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* (2002)31:1610.
- Miranda, M. J., and J. W. Glauber, "Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets," *American Journal of Agricultural Economics*, (2007)79-1: 206-215.
- National Agricultural Statistics Service, *QuickStats Database*, 2015, Accessed via *Ag-Analytics.Org*, December, 2015.
- 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, 2015.

Olson, K.R., J.M. Lang, J.D. Garcia Paredes, R.N. Majchrak, C.I. Hadley, M.E. Woolery, and R.M. Rejesus, *Illinois Bulletin 810, Average Crop, Pasture, and Forestry Productivity Ratings for Illinois Soils*, University of Illinois, 2000, available online at: <http://soilproductivity.nres.illinois.edu/Bulletin810ALL.pdf>

Peng, Shaobing, Jianliang Huang, John E. Sheehy, Rebecca C. Laza, Romeo M. Visperas, Xuhua Zhong, Grace S. Centeno, Gurdev S. Khush, and Kenneth G. Cassman, "Rice yields decline with higher night temperature from global warming," *Proceedings of the National Academy of Science*, (2004)101-27:997.

Popp, M., M. Rudstrom, and P. Manning. "Spatial Yield Risk Across Region, Crop, and Aggregation Method." *Can. J. Agr. Econ.* (2005)53:103-115

Risk Management Agency, Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures (by Josephson, G. R., R. B. Lord, and C. W. Mitchell Prepared for USDA–Risk Management Agency by Milliman & Robertson, Inc.), 2000.

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* (2002)57:499.

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.D., "Theme Overview: Current Issues in Risk Management and U.S. Agricultural Policy," *Choices* (Sept. 2013).

Woodard, J., "A Conditional Distribution Approach to Assessing the Impact of Weather Sample Heterogeneity on Yield Risk Estimation and Crop Insurance Ratemaking," *North American Actuarial Journal*. (2014)18-2:279-293

Woodard, J.D., "Data Science and Management for Large Scale Empirical Applications in Agricultural and Applied Economics Research," *Applied Economic Perspectives and Policy*, (2016) 38-3, pp. 373–388.

Woodard, J.D., "Ag-Analytics.org Data Warehouse", 2016, *Available online at: ag-analytics.org* (Accessed: December, 2015)

Woodard, J.D. "Big data and Ag-Analytics: an open source, open data platform for agricultural & environmental finance, insurance, and risk," *Agricultural Finance Review*, (May 2016)76-1:15-26,

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

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

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

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

Tables (1 per page)

Table 1: Farm yield panel regression results. Regression results of yield models, conditional on soil quality (Illinois Bulletin 810; Olson et al., 2000), and a county level time proxy for Illinois corn producers in the Farm Business Farm Management dataset. The dataset represents a panel (time series, cross section) dataset of farm level yields. TREND and INTERCEPT terms are shown at their means across counties.

Variables	Model 1	Model 2	Model 3	Model 4
<i>SOIL</i>	1.170*** (0.015)	1.147*** (0.015)	1.001*** (0.240)	0.962*** (0.235)
<i>SOIL</i> ²			0.001 (0.001)	0.001 (0.001)
<i>TIME</i>	1.184*** (0.155)	0.048 (0.587)	1.277*** (0.263)	0.048 (0.587)
<i>TIME</i> ²		-0.436*** (0.119)		-0.236 (0.147)
<i>INTERCEPT</i>	25.725*** (3.544)	-12.055** (5.740)	-16.661 (13.454)	-2.076 (14.351)
N	121,416	121,416	121,416	121,416
Adj. R ²	0.400	0.422	0.400	0.422
σ^2	747.917	720.873	747.920	720.876

Table 2. Yield risk (standard deviation) regression results: Regression results of yield risk models conditional on soil quality (Illinois Bulletin 810; Olson et al., 2000) for Illinois corn producers in the Farm Business Farm Management dataset. Standard deviation of farm yield was the independent variable, where standard deviation was estimated for each farm with at least 20 years of data. INTERCEPT are county fixed effects terms shown at their means.

Variables	Model 1	Model 2	Model 3
<i>SOIL</i>	-0.117*** <i>-0.013</i>	0.809*** <i>0.216</i>	0.966*** <i>0.218</i>
<i>SOIL</i> ²		-0.004*** <i>0.001</i>	-0.004*** <i>-0.001</i>
<i>AREA</i>			-0.00342*** <i>0.001</i>
<i>AREA</i> ²			0.000001*** <i>0.000</i>
<i>INTERCEPT</i>	37.802*** <i>-1.827</i>	-11.834 <i>11.714</i>	-11.834 <i>11.714</i>
N	3861	3861	3861
Adj. R ²	0.172	0.175	0.185
σ ²	19.222	19.135	18.921

Table 3: Regression of Soil Conditioned Rates on Actual USDA-RMA Rates. Table presents results of regression of soil conditioned rates on actual Federal Crop Insurance Program rate estimates from USDA-Risk Management Agency (RMA) methodology by Common Land Unit (CLU) for McLean County, Illinois corn. If RMA rates sufficiently differentiate rates by soil quality, then the parameter on the RMA Premium Rate Variable, should be not be significantly different from 1.0; a value of greater than 1.0 indicates RMA rates do not differentiate risks fully by soil type. The null is tested against whether the rate parameter equals 1.0. Standard errors are in parentheses, and significance is denoted by ***=1%, **=5%, *=10%

	<i>85% Coverage Level</i>	<i>65% Coverage Level</i>
<i>Intercept</i>	-0.06586*** (0.00007)	-0.02188*** (0.00005)
<i>RMA Premium Rate</i>	4.82580*** (0.00437)	2.61150*** (0.00622)
Adjusted R Square	0.99425	0.96150
Standard Error	0.00033	0.00025
Observations	7049	7049

Figures (1 per page)

Figure 1: Expected yield of corn by Common Land Unit field in McLean County, IL.

Heatmap of expected yield conditional on soil type from Model 1 for each Common Land Unit in McLean County, Illinois. Green indicates lower values while red units indicate higher values.

Figure 2: Kernel density of Soil Productivity Index of all Common Land Units in McLean County Illinois for Illinois Bulletin 810 (Olson, et al., 2000) Soil Productivity Rating

Figure 3: Differences in soil-naïve and soil-conditioned rates by Actual Production History. RMA rates (red curve) and soil-conditioned rates (blue curve) are displayed along APH for CLUs in McLean County, IL (n=7049). Results here are shown for 85% coverage level.