Title: Utilizing Large-Scale Insurance Datasets to Calibrate Sub-County Level Crop Yields

Products Authors: Francis Tsiboe, Dylan Turner, and Jisang Yu

Date: 10/02/2024

Note: The material contained herein is supplementary to the article named in the title and published in the Journal of Risk and Insurance.

Note S1: Details on Creating Validation Dataset

To transform the KFMA data set into something indicative of the SOBTPU, we first define insureds as the unique combinations of farm id, crop, irrigation practice, and crop year in the KFMA data. For the purposes of simulating aggregated level crop insurance outcomes like the SOBTPU, we formulate an insurance pool, whose designation is defined by unique combinations of USDA Crop Reporting District (CRD) (nine in KS), crop (corn, sorghum, soybeans, or winter wheat), insurance unit (OU, BU, or EU), insurance plan (Yield Protection [YP], Actual Production History [APH], Revenue Protection [RP], or RP with Harvest Price Exclusion [RP-HPE]), irrigation practice (dryland or irrigated), and coverage level (alternatives of 50-85% in 5% increments). The following outlines the steps involved in the simulation.

Following previous literature (Tsiboe & Tack, 2022), an algorithm based on RMA guidelines (Risk Management Agency [RMA], 2018) is used to estimate rate yield ($\bar{y}_{it}$) and approved yield ($\ddot{y}_{it}$) for each insured with their observed yields (y_{it}) in ten successive years serving as the basis for an APH database.[1] To maintain tractability of the simulation, we shed complexities such as yield exclusion, yield substitution, and trend adjustment in the calculations of both rate and approved yield. Where at least one yield observation is available, the rate yield is taken as the mean of the actual yields in the APH database. If there are no actual yields, the rate yield is taken as the T-yield retrieved from RMA's ADM. Approved yield calculations require actual or assigned yields for at least the last four successive years. Where the APH database has actual yields for the last four successive years, the approved yield is just the simple average of the actual yields. Where the APH database for the last four successive years is incomplete, missing entries are assigned a value based on variable T-yields to meet the four-year minimum yield requirement before taking the simple average. Missing entries in APH databases with zero, one, two, and three missing actual yields in the last four successive years are replaced with 65, 80, 90, and 100% of their counties T-yield, respectively. The simple average approved yield is then bounded between 90% of its value in the previous year (i.e., yield cap) and 70–80% of the relevant T-yield (i.e., yield floor). If the APH database has three, two, and one missing entry in the last four successive years, the floor is set at 70, 75, and 80%.

Crop insurance plans and coverage level are randomly assigned to each insured such that for each county/crop/irrigation practice combination, the distribution of policies follows the actual observed distributions in RMA's summary of business data. These randomly assigned policies are then linked to their respective crop year (2011-2022) ADM to retrieve policy level continuous rating parameters, subsidy factor, and contract prices (projected and harvest). Given the insurance plan, price election percent are randomly (with replacement) assigned to each insured's policy such that they fall within RMA's allowable ranges which we allow to change in 1% increment. Insured share is also randomly (with replacement) assigned to each insured's policy such that they fall within the range 10-100% in 5% increment. The planted area for each insured is taken as their observed value in the KFMA and rate of damage for a peril is taken as unity (i.e., reduction in yield from the guaranteed level is applied equally to both insured and uninsured acres). Finally, to calculate indemnities, the end-of-season market year average price is taken as the state level marketing year average price retrieved from NASS.

[1] Unlike (Tsiboe & Tack, 2022) who also estimate T-yield and reference yields from the KFMA data, this study instead takes these as given by RMA.

Crop insurance loss experience outcomes are calculated for each insured exactly as outlined in P21-1 and P21-2 of RMA's Appendix III/M-13 Handbook. The yield guarantee for each insured was taken as the product of their approved yield, projected price, and their randomly assigned coverage level and price election percent. The total liability is then defined as the product of approved yield and insured acres (i.e., planted acres multiplied by their randomly assigned insured share). Using each farms' calculated rate yield, randomly assigned policy, the continuous rating formula is then used to calculate the premium rate for each farm with the respective crop year policy environment serving as the basis for the rating parameters. The total premium was then taken as the liability multiplied by the premium rate. The subsidy amount for the policy is taken as the product of the total premium and the subsidy factor. The indemnity was then calculated as the difference between the yield guarantee and the observed yield if the latter is less than the former (recorded as zero otherwise) then multiplied by total acres insured.

A summary of business aggregated by insurance pools (as defined previously) is then generated from the simulated farm level outcomes by summing up insured acres, liability, total premium, and indemnity by insurance pool.

Table S1: Mean and standard deviation of farm level characteristics and simulated insurance experience for selected farms in the Kansas Farm Management Association (KFMA) from 1991 to 2022

	Wheat	Corn	Sorghum	Soybeans
Panel A: Actual farm enterprise level characteristics				
Number of dependents	2.823 (1.689)	2.888 (1.721)	2.837 (1.615)	2.809 (1.578)
Principal farm operator age (years)	53.732 (13.372)	53.216 (13.128)	52.824 (13.012)	54.061 (13.110)
Yield (bu/acre)	40.231 (16.112)	114.184 (53.292)	70.501 (29.750)	33.804 (16.014)
Harvested area (acres)	460.674 (488.175)	332.813 (486.141)	241.003 (265.832)	371.412 (457.391)
Number of operators	1.035 (0.466)	1.093 (0.530)	1.050 (0.444)	1.079 (0.514)
Number of workers	1.586 (1.383)	1.778 (1.617)	1.555 (1.318)	1.639 (1.474)
Labor (hours/acre)	0.330 (0.584)	0.345 (0.465)	0.339 (0.405)	0.376 (0.658)
Fertilizer ($/acre)	28.286 (28.577)	36.525 (33.308)	24.205 (20.570)	34.567 (30.720)
Rent ($/acre)	0.650 (0.310)	0.669 (0.300)	0.662 (0.302)	0.662 (0.305)
Capital ($/acre)	1952.815 (3350.950)	2408.850 (3671.571)	1738.210 (3588.813)	2529.496 (4363.702)
Other cost ($/acre)	63.311 (55.550)	74.336 (62.535)	56.800 (45.919)	73.304 (54.242)
Health insurance expenditure ($/dependent)	846.438 (1758.920)	974.741 (1917.863)	714.975 (1593.083)	819.520 (1649.529)
Life insurance expenditure ($/dependent)	392.467 (1421.832)	484.736 (1965.754)	335.725 (1485.742)	375.552 (1799.064)
Number of observations	45,491	35,019	30,496	37,237
Panel B: Simulated pseudo insurance pool level characteristics				
Net insured area (acres)	747.394 (1156.077)	621.641 (1020.146)	355.599 (480.390)	763.377 (1176.995)
Total liability ($/acre)	210.476 (67.929)	508.869 (203.355)	262.814 (87.961)	364.193 (136.389)
Total premium ($/acre)	33.079 (15.676)	65.379 (30.751)	47.871 (22.796)	47.832 (27.333)
Total subsidy ($/acre)	20.636 (9.167)	40.757 (18.169)	30.339 (13.935)	29.743 (16.071)
Total indemnity ($/acre)	11.532 (32.447)	41.546 (104.282)	34.704 (75.075)	25.357 (67.223)
Premium per dollar of liability	0.160 (0.078)	0.144 (0.079)	0.192 (0.102)	0.144 (0.093)
Subsidy per dollar of premium	0.635 (0.051)	0.634 (0.054)	0.641 (0.046)	0.634 (0.052)
Number of observations	3,615	4,606	2,615	4,220

Table S2: Regression results - All listed crop and wheat

	All listed crops			Wheat		
	Mean of end of season yield	CV of end of season yield	Lost cost ratio	Mean of end of season yield	CV of end of season yield	Lost cost ratio
Unit structure / Farm bill period categories [base = OU in 2000]						
BU in 2000	-0.067*** (0.007)	0.027*** (0.004)	-1.346*** (0.041)	-0.094*** (0.019)	0.044*** (0.014)	-1.338*** (0.116)
EU/WU in 2000	0.508*** (0.008)	-0.286*** (0.013)	-4.626*** (0.091)	0.542*** (0.107)	0.022 (0.129)	-5.347*** (0.887)
OU in 2008	0.585*** (0.007)	-0.300*** (0.006)	-1.272*** (0.050)	0.439*** (0.021)	-0.231*** (0.027)	-3.369*** (0.139)
BU in 2008	0.515*** (0.008)	-0.220*** (0.006)	-3.076*** (0.055)	0.351*** (0.026)	-0.166*** (0.028)	-5.472*** (0.155)
EU/WU in 2008	0.913*** (0.008)	-0.379*** (0.007)	-3.706*** (0.064)	0.780*** (0.034)	-0.329*** (0.034)	-7.534*** (0.199)
OU in 2014	0.224*** (0.011)	-0.440*** (0.006)	-1.777*** (0.068)	0.332*** (0.028)	-0.367*** (0.022)	-2.432*** (0.177)
BU in 2014	0.196*** (0.011)	-0.312*** (0.006)	-3.742*** (0.068)	0.321*** (0.031)	-0.289*** (0.021)	-4.135*** (0.181)
EU/WU in 2014	0.524*** (0.010)	-0.489*** (0.006)	-4.055*** (0.071)	0.695*** (0.028)	-0.444*** (0.023)	-5.917*** (0.194)
OU in 2018	-0.069*** (0.015)	-0.150*** (0.006)	-0.299*** (0.084)	0.279*** (0.037)	-0.016 (0.021)	-1.593*** (0.211)
BU in 2018	-0.135*** (0.016)	-0.035*** (0.005)	-1.906*** (0.084)	0.203*** (0.038)	0.097*** (0.021)	-2.610*** (0.214)
EU/WU in 2018	0.315*** (0.014)	-0.206*** (0.005)	-2.078*** (0.086)	0.656*** (0.037)	-0.077*** (0.022)	-4.130*** (0.222)
Precipitation	0.064*** (0.004)	-0.027*** (0.003)	-0.354*** (0.012)	0.067*** (0.010)	-0.030*** (0.007)	-0.816*** (0.030)
Precipitation Squared	-0.003*** (0.000)	0.001*** (0.000)	0.018*** (0.001)	-0.003*** (0.001)	0.001*** (0.000)	0.045*** (0.002)
Precipitation Cubed	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000 (0.000)	-0.001*** (0.000)
Degree Days: 0-10C	0.000*** (0.000)	0.000*** (0.000)	0.002*** (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Degree Days: 10-30C	0.001*** (0.000)	0.000*** (0.000)	-0.007*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	-0.004*** (0.000)
Degree Days: +30C	-0.008*** (0.000)	0.003*** (0.000)	0.091*** (0.001)	-0.007*** (0.000)	0.002*** (0.000)	0.059*** (0.002)
Projected commodity price	-0.913*** (0.005)	0.080*** (0.004)	1.649*** (0.030)	-0.410*** (0.026)	0.274*** (0.033)	4.644*** (0.122)
County base rate	-0.498*** (0.003)	0.167*** (0.002)	1.087*** (0.017)	-0.653*** (0.010)	0.226*** (0.006)	2.227*** (0.049)
Trend	-0.055*** (0.001)	-	0.816*** (0.008)	-0.075*** (0.004)	-	1.001*** (0.022)
Trend Squared	0.003*** (0.000)	-	-0.034*** (0.000)	0.003*** (0.000)	-	-0.038*** (0.001)
Constant	3.385*** (0.027)	-0.270*** (0.021)	-11.160*** (0.105)	2.023*** (0.064)	-0.498*** (0.061)	-11.969*** (0.271)
Number of observations	1544747	349360	1544747	244508	58233	244508
Number of insurance pools	240480	170065	240480	42983	29543	42983
R-squared within model	0.005	0.062	0.046	0.003	0.084	0.034
R-squared between model	0.245	0.124	0.039	0.142	0.091	0.076
R-squared overall model	0.137	0.092	0.039	0.077	0.086	0.054

Significance levels - *p<0.1 ** p<0.05, ***p<0.01.

Standard errors in parentheses are clustered by insurance pool.

Table S3: Regression results - Corn and Soybeans

	Corn			Soybeans		
	Mean of end of season yield	CV of end of season yield	Lost cost ratio	Mean of end of season yield	CV of end of season yield	Lost cost ratio
Unit structure / Farm bill period categories [base = OU in 2000]						
BU in 2000	-0.097*** (0.011)	0.027*** (0.006)	-1.259*** (0.061)	-0.043*** (0.009)	0.013** (0.006)	-1.498*** (0.062)
EU/WU in 2000	0.458*** (0.010)	-0.274*** (0.018)	-4.324*** (0.127)	0.449*** (0.009)	-0.339*** (0.018)	-4.939*** (0.131)
OU in 2008	0.329*** (0.010)	-0.225*** (0.009)	-1.136*** (0.073)	0.331*** (0.011)	-0.578*** (0.023)	-1.397*** (0.085)
BU in 2008	0.225*** (0.013)	-0.123*** (0.009)	-2.757*** (0.079)	0.300*** (0.012)	-0.511*** (0.023)	-3.418*** (0.091)
EU/WU in 2008	0.626*** (0.012)	-0.280*** (0.011)	-2.880*** (0.091)	0.644*** (0.012)	-0.648*** (0.026)	-4.259*** (0.103)
OU in 2014	0.153*** (0.016)	-0.570*** (0.010)	-3.602*** (0.101)	0.415*** (0.016)	-0.415*** (0.019)	0.374*** (0.108)
BU in 2014	0.129*** (0.017)	-0.403*** (0.009)	-5.353*** (0.100)	0.371*** (0.017)	-0.292*** (0.018)	-1.936*** (0.107)
EU/WU in 2014	0.429*** (0.015)	-0.593*** (0.010)	-5.503*** (0.105)	0.658*** (0.015)	-0.485*** (0.018)	-1.862*** (0.111)
OU in 2018	-0.171*** (0.023)	-0.149*** (0.010)	-3.247*** (0.126)	0.321*** (0.023)	-0.300*** (0.021)	3.306*** (0.134)
BU in 2018	-0.226*** (0.024)	-0.013 (0.009)	-4.694*** (0.126)	0.254*** (0.024)	-0.203*** (0.020)	1.163*** (0.134)
EU/WU in 2018	0.211*** (0.021)	-0.181*** (0.009)	-4.696*** (0.128)	0.636*** (0.021)	-0.370*** (0.020)	1.419*** (0.136)
Precipitation	0.061*** (0.007)	-0.008 (0.005)	-0.234*** (0.021)	0.009 (0.008)	-0.030*** (0.007)	-0.738*** (0.022)
Precipitation Squared	-0.003*** (0.000)	0.000 (0.000)	0.015*** (0.001)	0.000 (0.000)	0.002*** (0.000)	0.032*** (0.001)
Precipitation Cubed	0.000*** (0.000)	0.000 (0.000)	0.000*** (0.000)	0.000 (0.000)	0.000*** (0.000)	0.000*** (0.000)
Degree Days: 0-10C	-0.001*** (0.000)	0.000*** (0.000)	0.004*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.004*** (0.000)
Degree Days: 10-30C	0.001*** (0.000)	0.000*** (0.000)	-0.011*** (0.000)	0.000*** (0.000)	0.000** (0.000)	-0.008*** (0.000)
Degree Days: +30C	-0.014*** (0.000)	0.004*** (0.000)	0.105*** (0.001)	-0.009*** (0.000)	0.004*** (0.000)	0.113*** (0.001)
Projected commodity price	-0.396*** (0.010)	0.088*** (0.009)	-0.925*** (0.047)	0.078*** (0.014)	0.188*** (0.030)	3.976*** (0.085)
County base rate	-0.445*** (0.005)	0.181*** (0.003)	0.890*** (0.026)	-0.407*** (0.004)	0.135*** (0.003)	1.182*** (0.028)
Trend	-0.072*** (0.002)	-	1.058*** (0.011)	-0.112*** (0.002)	-	0.493*** (0.013)
Trend Squared	0.004*** (0.000)	-	-0.034*** (0.001)	0.004*** (0.000)	-	-0.031*** (0.001)
Constant	3.279*** (0.044)	-0.428*** (0.037)	-9.411*** (0.168)	2.093*** (0.055)	-0.515*** (0.074)	-12.154*** (0.227)
Number of observations	691762	157001	691762	608477	134126	608477
Number of insurance pools	103046	73989	103046	94451	66533	94451
R-squared within model	0.014	0.123	0.060	0.010	0.044	0.048
R-squared between model	0.173	0.137	0.035	0.143	0.147	0.055
R-squared overall model	0.090	0.134	0.044	0.070	0.088	0.048

Significance levels - *p<0.1 ** p<0.05, ***p<0.01.

Standard errors in parentheses are clustered by insurance pool.

Figure S1: Percentage of time n_t (x-axis) and n_{t-1} (y-axis) were selected by solving equation (7)

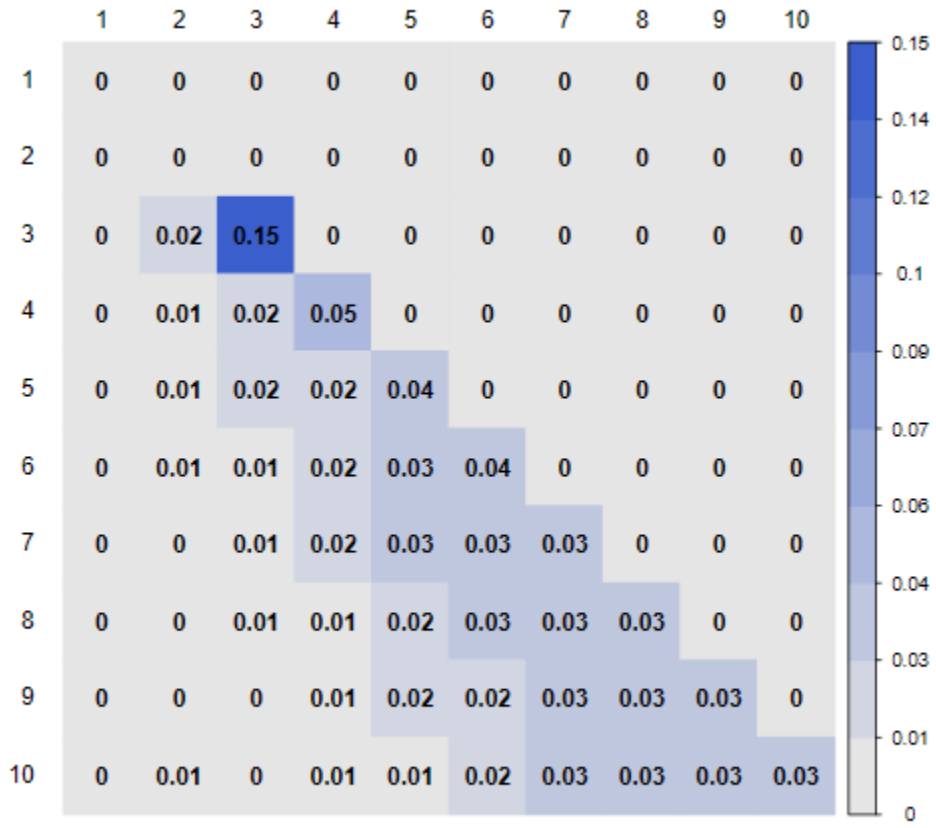

Figure S2: Distributions of calibrated corn yields disaggregated by SOBTPU characteristics

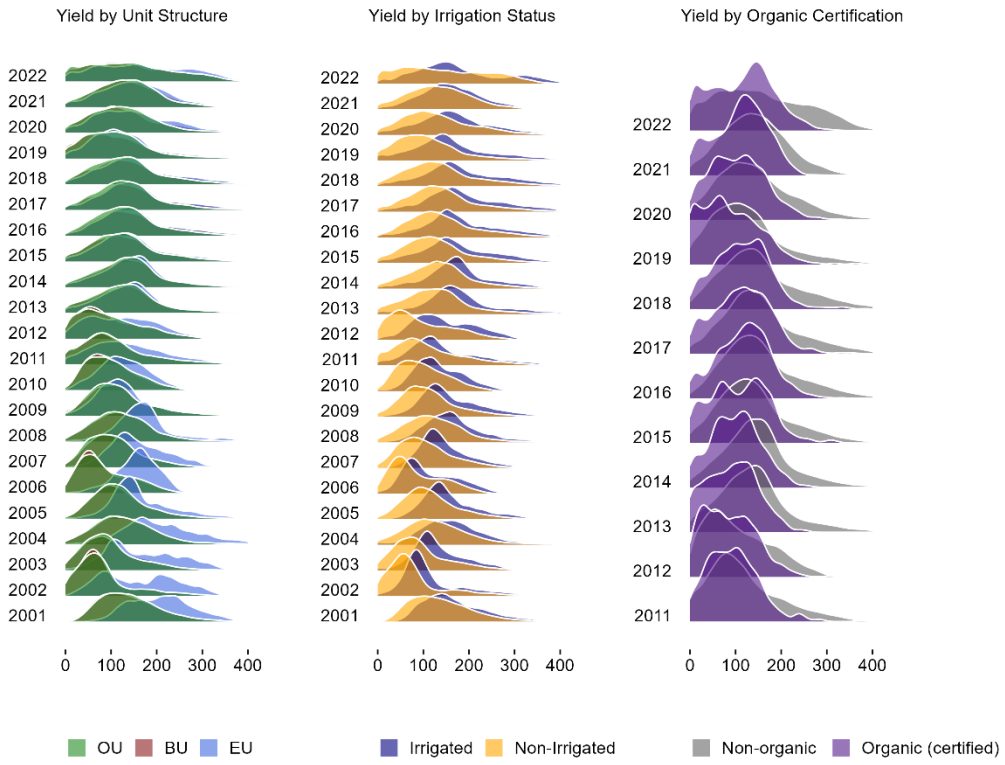

Figure S3: Distributions of calibrated soybean yields disaggregated by SOBTPU characteristics.

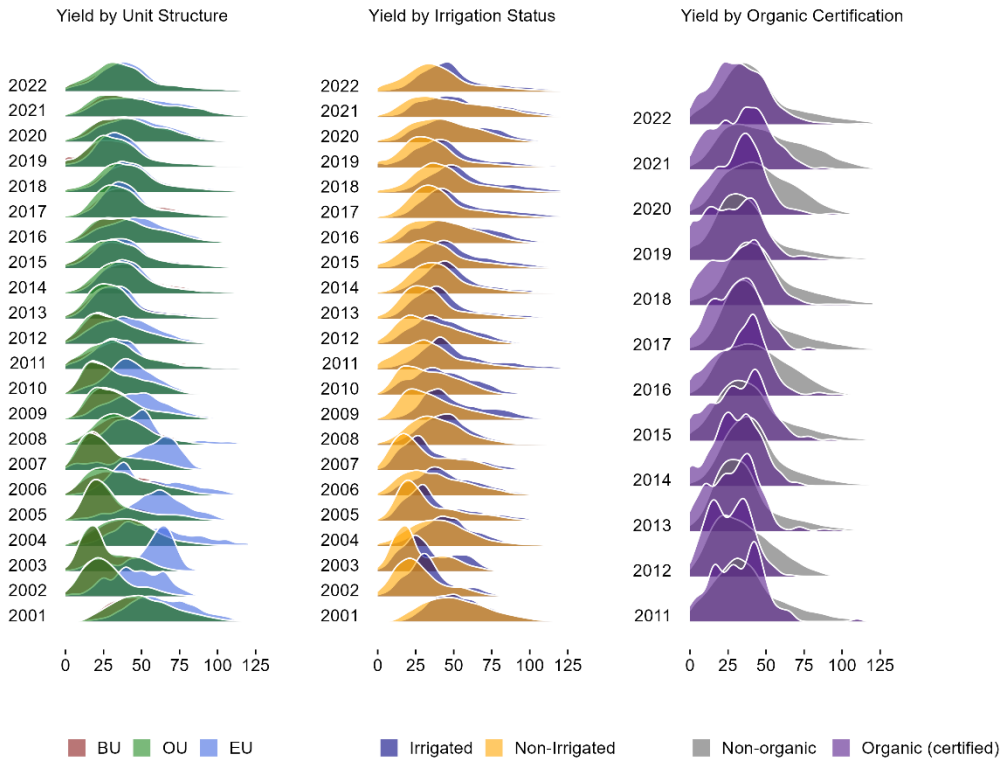

Figure S4: Distributions of calibrated wheat yields disaggregated by SOBTPU characteristics

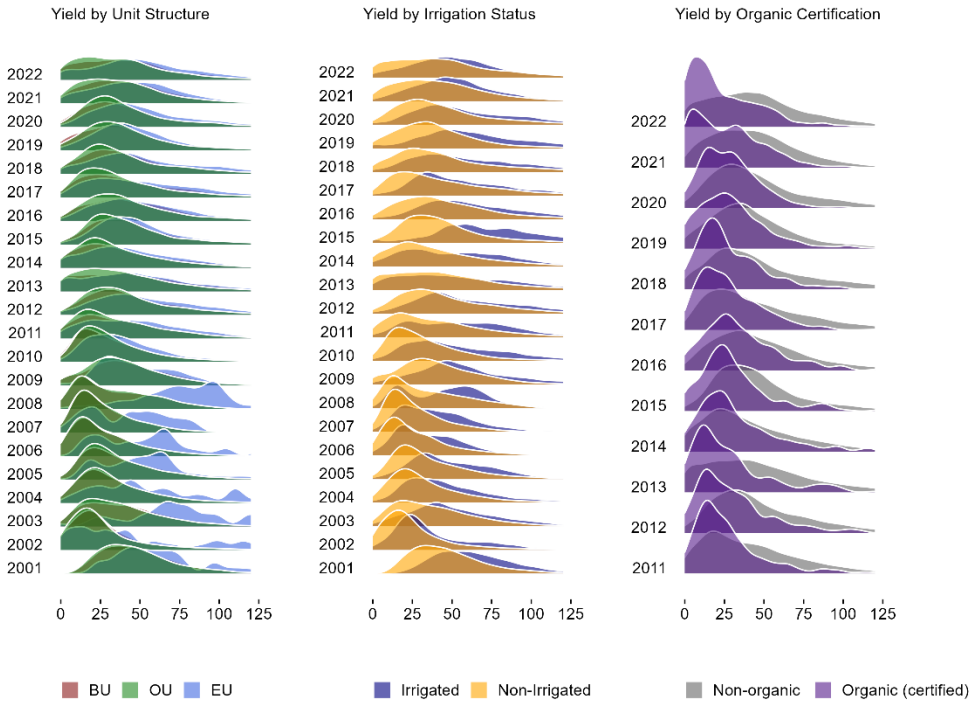

Figure S5: Boxplot of calibrated to actual yield ratios from simulated Kansas crop insurance data.

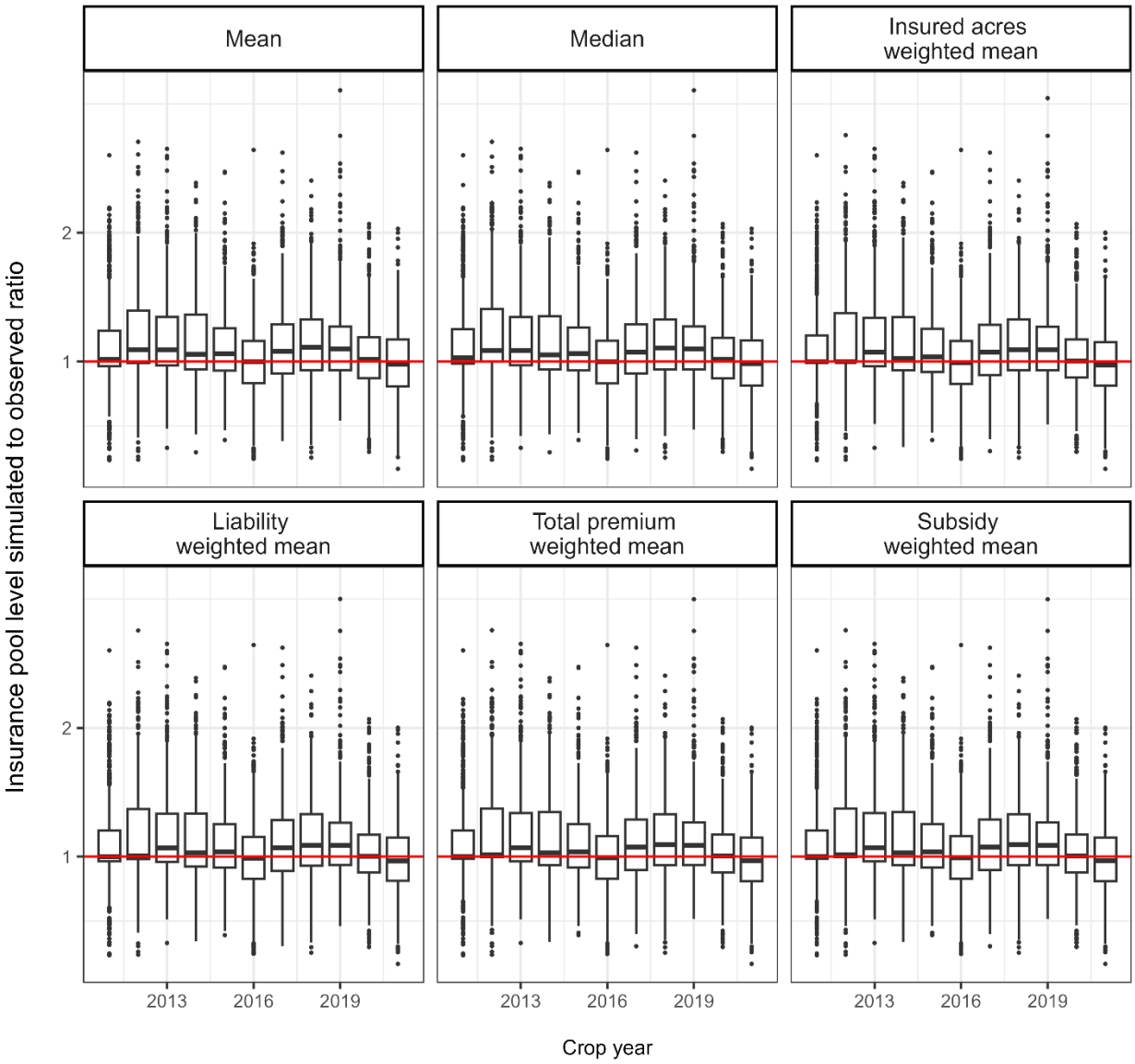

Figure S6: Effect of entry and exit yield on measurement error.

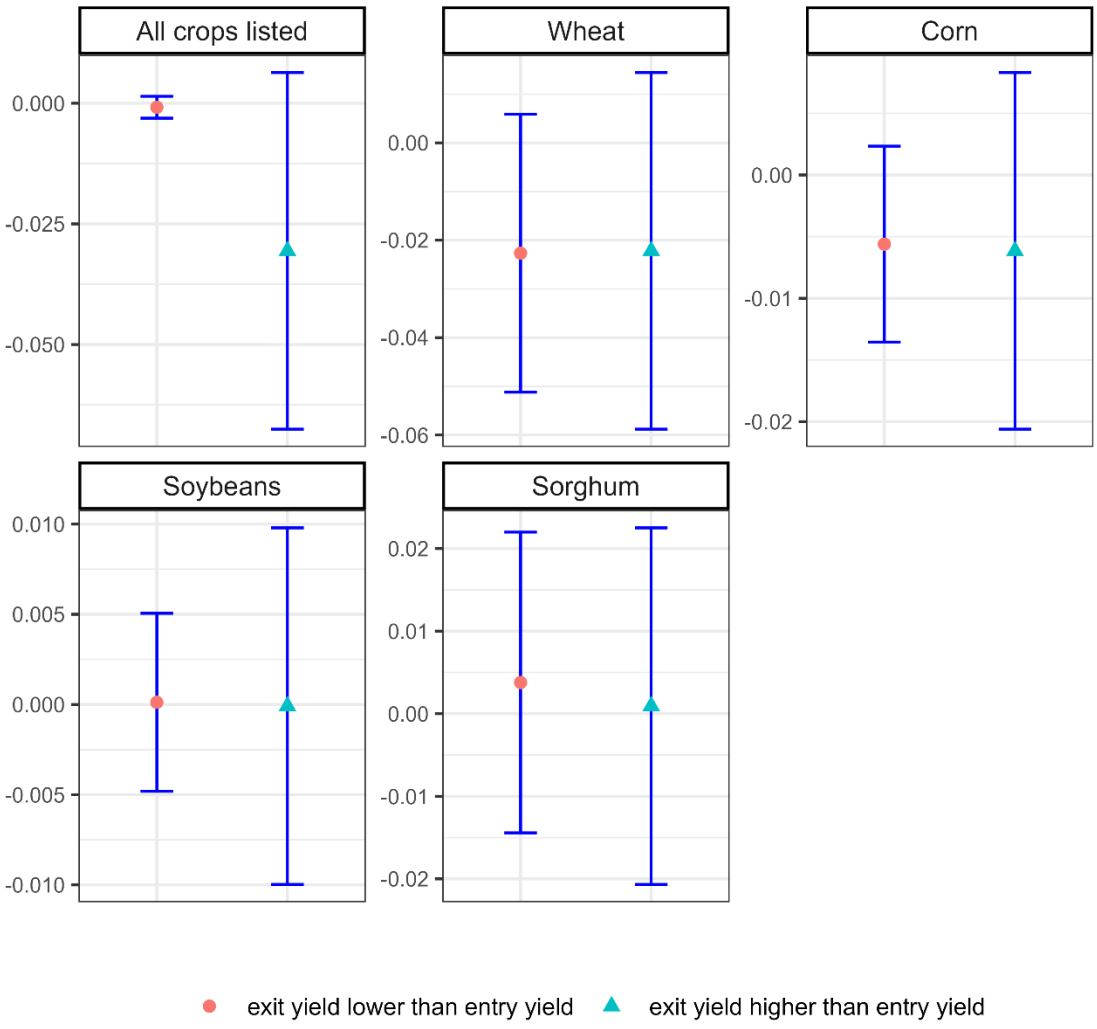

Figure S7: Histogram showing the difference between entering and exiting yield.