Supporting Information

Larsen 10.1073/pnas.1301900110

SI Text

Statistical Models. The cross-sectional approach, which uses one time period, attempts to directly quantify the effect of landscape simplification on insecticide use using a multiple linear regression. The advantage of this approach is that it accounts for substitution behavior of the farmers when pest problems occur. However, the cross-sectional approach is only valid if the effect of landscape simplification on insecticide use is consistently estimated (1). Unmeasured or unobserved characteristics, such as soil quality, are important determinants of land-use patterns in agricultural settings (2). Thus, the cross-sectional approach may confound landscape simplification with other unobserved characteristics. Because the omitted variable is, as the name suggests, omitted from the regression, the magnitude and sign of the omitted variable bias is generally unknown (1).

The advantage of the fixed-effects approach is that it removes the time-invariant unobserved effects unique to individual counties that may otherwise bias the estimates derived from cross-sectional analyses (1). Rather, the fixed-effects model exploits the year-to-year (census-to-census) variation in land use at the county level to estimate how landscape simplification affects insecticide use. Analysis of within-county variation in cropland revealed sufficient year-to-year variation for the fixed-effects estimation to be reasonably statistically precise.

Mathematically, the cross-sectional analysis is represented by the following equation:

$$Y_i = X_i'\beta + u_i,$$

where Y_i is the response variable, the proportion of harvested cropland sprayed with insecticides in county i. X represents the vector of covariates, including proportion of county in cropland, income per harvested hectare, and proportion of cropland in corn, soybeans and small grains, vegetables, and fruit and nut orchards in each county i, and u represents the random error term for each county, i. This model is analyzed for each US Department of Agriculture (USDA) Census of Agriculture (3) year using data from the Census, and from the National Agriculture Statistics Service Cropland Data Layer (4) for the 2007 specification.

The fixed-effects estimation is represented by the following equation:

$$Y_{it} = X_{it}'\beta + c_i + u_{it}$$

for county i in year t. Again, Y_{it} is the proportion of cropland treated with insecticides and X is a vector of covariates as described above, indexed for each time period t. c is the unobserved effects (or individual heterogeneity) term for each county i, that is assumed roughly time invariant over the period of analysis (1). By de-meaning the data, this time-invariant unobserved effect drops out of the above equation and the model is identified from the variation of observations for a given county, i, away from that county's mean. If the model also includes year fixed effects, then the model is identified from year-to-year variation within a county, after controlling for time trends shared by all counties in the study. If the model is estimated by assuming c_i is uncorrelated with the observable covariates (i.e., a random-effects model or pooled ordinary least-squares model), c_i is effectively put in the error term. Thus, if c_i is in fact correlated with the observed covariates (e.g., if unobserved soil quality is correlated with the observed proportion of county in cropland), the coefficient on proportion of county in cropland and other coefficients in the model will be inconsistent (1).

Insecticide Use Trends. I used the NASS Agricultural Pesticide Use Database (APUD; ref. 5) to check that proportion of cropland treated with insecticide from the Census of Agriculture reflected the amount of insecticides applied. The APUD is available for participating states, reporting a limited number of crops. I chose to examine the relationship between acres treated and active ingredients (active ingredients per acre) applied to corn because: (*i*) corn is reported in the APUD for most of the states in this study for two of the most recently available census years (1997, 2002); (*ii*) it represents a large proportion of the cropland in the Midwest; and (*iii*) other crops, such as soybeans, are not reported for most of the Midwest for both 1997 and 2002. Only state-level data are available and, thus, I assumed that state-level trends hold at the county level.

I find a strong linear relationship between acres treated and active ingredients per acre for both years (Fig. S2). The slope of the relationship is different between years, possibly reflecting differences in the rate (pound per acre per year) of insecticides available. For example, in Iowa in 1997 chlorpyrifos was applied at a rate of 1.730 pounds per acre per year to 6% of corn, whereas cyfluthrin was applied at a rate of 0.006 pounds per acre per year to 3% of corn. In 1992 chlorpyrifos was applied to 9% of corn and cyfluthrin was not used. Despite changes in concentrations or chemicals in 1997 compared with 2002, it is clear that there is a linear relationship between acres treated and active ingredients used in each year.

1. Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA), 1st Ed.
2. Deschênes O, Greenstone M (2007) The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *Am Econ Rev* 97(1):354–385.
3. USDA National Agricultural Statistics Service (2012) *Census of Agriculture*. Available at http://www.agcensus.usda.gov/index.php. Accessed May 20, 2012.
4. USDA National Agricultural Statistics Service (2012) Cropland Data Layer. Available at http://nassgeodata.gmu.edu/CropScape/. Accessed May 20, 2012.
5. USDA National Agricultural Statistics Service (2012) Agricultural Chemical Use Database. Available at www.pestmanagement.info/nass. Accessed September 1, 2012.

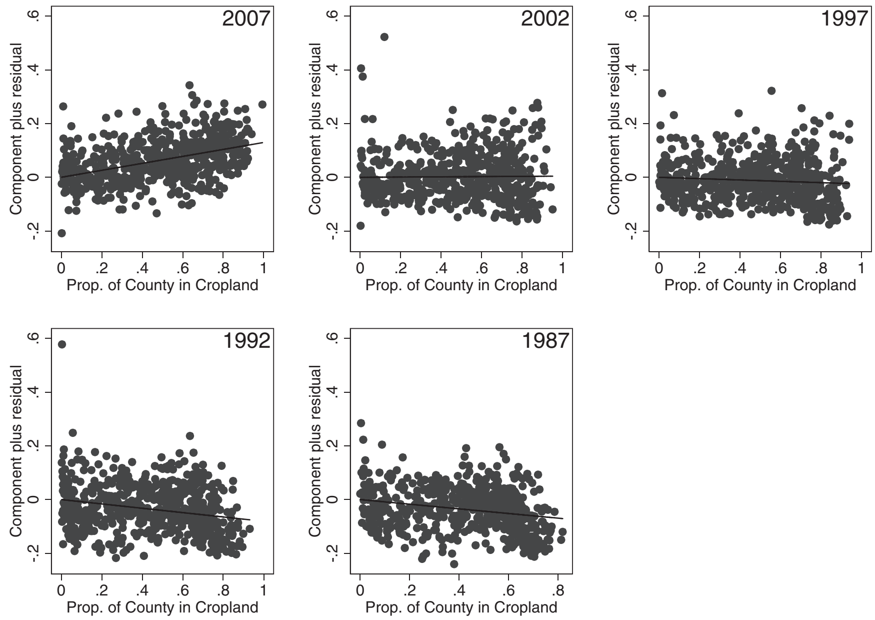

Fig. S1. Partial residual plots of the effect of proportion of the county in cropland on proportion of cropland treated with insecticides over time from 2007 to 1987, including all counties. The coefficient on proportion of county in cropland (i.e., land simplification) is not a consistent driver of insecticide use. The number of points plotted in each year plot corresponds to the number of observations in Table 2. The slopes and SEs for proportion of county in cropland for each year also correspond to Table 2.

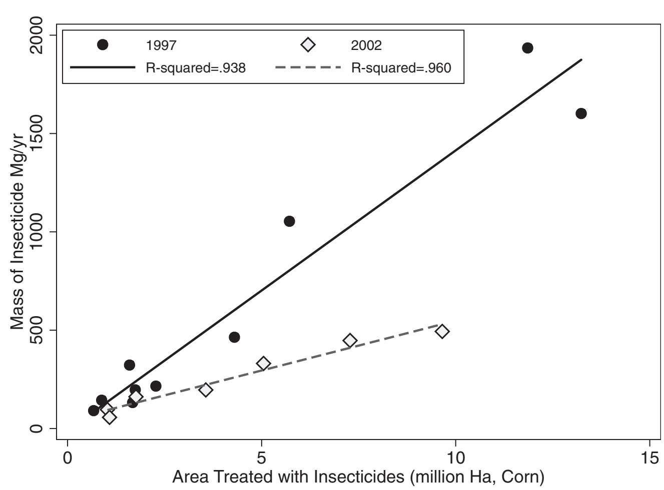

Fig. S2. Area treated with insecticides is linearly related to mass of insecticides applied at the state level. Points represent individual state data for 1997 (●) and 2002 (◊) with data for insecticides use in corn from the NASS APUD. The difference in slope is likely a result of different rates and different levels of application between years (*SI Text*).

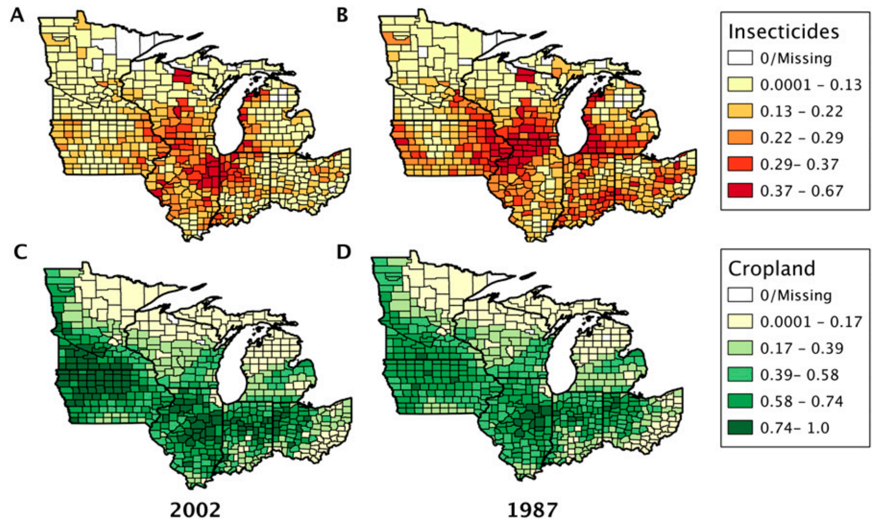

Fig. S3. Map of proportion of cropland treated with insecticides (*A* and *B*) and proportion of the county in cropland (*C* and *D*) for 2002 and 1987. The proportion of cropland treated with insecticides varies greatly between years both within the same county and between counties. The legend is based on 2007 quintiles with the range of the first and last quintile extended to incorporate the lowest (highest) values of the other years.

Table S1. Evaluating the endogeneity of income

	2007 Model specifications			2007 Model specifications excluding income		
Model components	1	2	3	1	2	3
Proportion of county in cropland	0.0975**	0.1362**	0.1297**	0.0989**	0.1408**	0.1353**
	(0.0343)	(0.0293)	(0.0296)	(0.0348)	(0.0296)	(0.0302)
Income per harvested hectare	0.0001	0.0002**	0.0001*			
	(0.0001)	(0.0000)	(0.0000)			
Proportion corn	0.4957**	0.4008**	0.4100**	0.5027**	0.4122**	0.4115**
	(0.0962)	(0.0697)	(0.0671)	(0.0926)	(0.0701)	(0.0681)
Proportion soybeans and small grains	0.0508	−0.0029	−0.0000	0.0472	−0.0121	−0.0078
	(0.0386)	(0.0356)	(0.0367)	(0.0376)	(0.0363)	(0.0380)
Proportion fruit and vegetables	0.9549**	0.8928**	0.8546**	0.9612**	0.9121**	0.9102**
	(0.0790)	(0.0545)	(0.0631)	(0.0802)	(0.0605)	(0.0610)
Constant	−0.0450	−0.0134	−0.0095	−0.0406	−0.0031	−0.0010
	(0.0335)	(0.0147)	(0.0139)	(0.0344)	(0.0162)	(0.0156)
Observations	562	596	603	562	596	603
R^2	0.59	0.65	0.65	0.59	0.64	0.64
SE clusters	ASD	ASD	ASD	ASD	ASD	ASD
No. clusters	62	63	63	62	63	63
NASS cropland	X			X		
Census cropland		X	X		X	X
NASS selection	X			X		
Census selection		X			X	

Column descriptions apply to both model specifications including and not including income. In column 1 cropland was defined based on the NASS CDL (*SI Text*). In column 2 cropland was defined as total harvested acres based on the Census of Agriculture. Counties were excluded if the proportion of county in cropland was less than 0.01. Column 3 maintained the census definition of cropland but did not exclude any counties regardless of the proportion of the county in cropland. The right half of the table excludes the income covariate to evaluate potential endogeneity between income and insecticide use. Cluster-robust SEs were used for all models, clustering on the agricultural statistical district within a state for single-year models and clustering on the county for multiyear models. X indicates the data source and selection criteria. For each covariate, the regression coefficient is the top number in the cell with the standard errors below in parentheses. *$P < 0.05$, **$P < 0.01$.

Table S2. Cross-sectional analysis of census years 2007–1987 for NASS and Census specifications

	NASS					Census				
Model components	2007	2002	1997	1992	1987	2007	2002	1997	1992	1987
Proportion of county in cropland	0.1176**	−0.0134	−0.0320	−0.1151**	−0.0653	0.1362**	−0.0024	−0.0282	−0.1012**	−0.0847*
	(0.0367)	(0.0452)	(0.0411)	(0.0392)	(0.0372)	(0.0293)	(0.0444)	(0.0403)	(0.0377)	(0.0372)
Income per harvested hectare	0.0001	−0.0002	−0.0001	0.0000	−0.0002	0.0002**	0.0000	0.0002*	0.0003	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Proportion corn	0.4612**	0.4073**	0.3907**	0.5158**	0.6852**	0.4008**	0.3533**	0.3350**	0.4263**	0.6167**
	(0.1014)	(0.0732)	(0.0860)	(0.0970)	(0.0944)	(0.0697)	(0.0544)	(0.0655)	(0.0747)	(0.0721)
Proportion soybeans and small grains	0.0331	0.0206	−0.0184	0.0025	0.0100	−0.0029	0.0125	−0.0239	−0.0364	0.0128
	(0.0393)	(0.0415)	(0.0411)	(0.0448)	(0.0463)	(0.0356)	(0.0358)	(0.0390)	(0.0472)	(0.0442)
Proportion fruit and vegetables	0.9515**	0.9486**	0.9116**	0.8053**	0.8843**	0.8928**	0.8853**	0.8419**	0.7404**	0.8484**
	(0.0795)	(0.0863)	(0.0754)	(0.0971)	(0.0741)	(0.0545)	(0.0656)	(0.0586)	(0.0748)	(0.0657)
Constant	−0.0404	−0.0052	0.0070	−0.0052	−0.0366	−0.0134	0.0038	0.0096	0.0285	−0.0256
	(0.0337)	(0.0275)	(0.0300)	(0.0337)	(0.0323)	(0.0147)	(0.0142)	(0.0145)	(0.0244)	(0.0204)
Observations	562	563	565	566	564	596	596	603	603	600
R^2	0.59	0.31	0.33	0.41	0.54	0.65	0.32	0.37	0.41	0.56
SE clusters	ASD	ASD	ASD	ASD	ASD	ASD	ASD	ASD	ASD	ASD
No. clusters	62	62	62	62	62	63	62	63	63	63

For all specifications cropland was defined using the census metric, total harvested acres. In the NASS section, counties were included if they met the NASS selection criteria (NASS proportion cropland >0.01) in 2007. In the Census section, counties were included if they met the census selection criteria. The number of counties in each year varies because of counties missing data on insecticides or cropland were dropped from the regression in that year. For both specifications, the relationship between proportion of county in cropland and proportion of cropland treated with insecticides is absent or reversed compared with the 2007 results. For each covariate, the regression coefficient is the top number in the cell with the standard errors below in parentheses. $*P < 0.05$, $**P < 0.01$.

Table S3. Fixed effects analysis for different model specifications with, year, county, and year and county fixed effects

Model components	1	2	3	4	5	6	7	8	9
Proportion of county in cropland	−0.0212	−0.0691	−0.0126	−0.0136	−0.0967*	−0.0079	−0.0164	−0.1021*	−0.0089
	(0.0171)	(0.0455)	(0.0411)	(0.0165)	(0.0434)	(0.0400)	(0.0169)	(0.0420)	(0.0401)
Income per harvested hectare	−0.0001	0.0004**	0.0000	0.0002*	0.0003**	0.0000	0.0001	0.0003**	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Proportion corn	0.5041**	0.6447**	0.4126**	0.4411**	0.6718**	0.4782**	0.4484**	0.6844**	0.4764**
	(0.0378)	(0.0530)	(0.0648)	(0.0300)	(0.0494)	(0.0591)	(0.0323)	(0.0488)	(0.0585)
Proportion soybeans and small grains	−0.0032	−0.3531**	−0.4522**	−0.0234	−0.3464**	−0.3984**	−0.0256	−0.3401**	−0.4037**
	(0.0200)	(0.0501)	(0.0546)	(0.0177)	(0.0479)	(0.0513)	(0.0185)	(0.0484)	(0.0513)
Proportion fruit and vegetables	0.8876**	0.4571**	0.4152**	0.8302**	0.3698**	0.3022*	0.8122**	0.4013**	0.3069**
	(0.0460)	(0.1051)	(0.1126)	(0.0368)	(0.1024)	(0.1188)	(0.0435)	(0.0850)	(0.0969)
Constant	−0.0160	0.0580	0.2192**	0.0017	0.0659	0.1641**	0.0068	0.0607	0.1683**
	(0.0150)	(0.0434)	(0.0494)	(0.0091)	(0.0392)	(0.0432)	(0.0113)	(0.0400)	(0.0446)
Observations	2,820	2,820	2,820	2,998	2,998	2,998	2,930	2,930	2,930
R^2	0.49	0.74	0.79	0.50	0.75	0.80	0.49	0.74	0.79
No. clusters	567	567	567	610	610	610	586	586	586
Year effects	Y	N	Y	Y	N	Y	Y	N	Y
County effects	N	Y	Y	N	Y	Y	N	Y	Y
Selection rule	NASS	NASS	NASS	Census	Census	Census	All, balanced	All, balanced	All, balanced

For columns 1–3 counties were included if they met the NASS section criteria in 2007. In columns 4–6 counties were included if they met the census selection criteria. I repeated the analysis including all counties that had observations in all time periods, creating a balanced panel (7–9). Year and county fixed effects are indicated for each column by Y (included) and N (not included). All models used cluster robust SEs, clustering on individual counties. For each covariate, the regression coefficient is the top number in the cell with the standard errors below in parentheses. $*P < 0.05$, $**P < 0.01$.