Appendix: Estimating the Link Between Payments and Concentration Growth

The study estimated the link between payments and concentration growth in several ways, both with and without controls. Results were similar in all cases. This appendix provides more detail about the methods used to derive results in tables 6 and 7 and figures 6 and 7, as well as some additional results obtained using different estimation procedures.

For the estimates reported in table 6, concentration is defined as the weighted-median cropland area in each ZIP Code (the acres of cropland on the farm for which half the cropland acreage occurs on farms with more cropland and half on farms with less). For each ZIP Code and panel, the percentage change in concentration is calculated as 100 times the change in concentration divided by average concentration in the two years considered. The long panels calculate the percentage change between 1987 and 2002. Payments per acre in each ZIP Code are calculated by dividing total payments in the beginning year by total cropland. ZIP Codes with no payments comprised one payment group and those with positive payments were classified into five quintiles, each with the same number of ZIP Codes. The payment-per-acre cutoffs are different in the different panels because payments vary across years. The cutoffs were chosen so as to have the same number of ZIP Codes in each quintile with positive payments. For the long panels, the payment groups were constructed using payments per acre in 1987. Because ZIP Codes receive different payment levels in different years and are sometimes classified into different payment groups in different panels, the percentage change for the long panel does not equal the sum of the individual panels.

The sample of ZIP Codes in the analysis is somewhat less than the population of ZIP Codes in the census of agriculture. ZIP Codes were dropped from the analysis if (1) less than three farms returned census forms in any of the four censuses examined; (2) if the ZIP Code reported no cropland in any of the panel years; (3) if commodity program payments per acre, the ratio of cropland area to ZIP Code area, or the ratio of crop sales to cropland area were in the top 2 percent of all ZIP Codes. The extreme observations were omitted from the analysis due to the highly skewed distributions of these variables. For example, some ZIP Codes may have extremely high payment-to-cropland ratios if there are very few acres of cropland, but a modest level of payments, perhaps as a result of historical plantings. After omitting these observations, the analyses include 21,524 ZIP Codes.

Results reported in table 7 (with controls) are derived from estimation of *generalized additive models* with the form:

$$\Delta c_i = X_i\beta + f(x_i,y_i) + g_c(c_{0i}) + g_a(a_{0i}) + g_s(s_{0i}) + \varepsilon_i$$

where subscript i (omitted below to simplify notation) indexes ZIP Codes, c_0 denotes concentration in the beginning year, Δ_c is the percent change in concentration ($(c_1 - c_0)/ \frac{1}{2}(c_1 + c_0)$), X is a matrix of indicator variables denoting payment-per-acre categories (one column of each row equals 1 and the others equal 0), β is a vector of payment-category effects, $f(x,y)$ is a smooth function of locations (x, y) of ZIP Code centroids, $g_c(c_0)$ is a smooth function of beginning-year concentration ($c0$), $g_a(a_0)$ is a smooth function of the ratio of cropland area to ZIP Code area in the beginning year (a_0), $g_s(s_0)$ is a smooth function of crop sales per acre of cropland (s_0), and ε is a random error. For the long panels, Δc is the sum of percentage changes in concentration in the three individual panels, and c_0, a_0, and s_0 are the average of values from 1987, 1992, and 1997.

The critical assumption is that, conditional on controls for location, beginning-year concentration, beginning-year ratio of cropland to ZIP Code area, and beginning-year crop sales per cropland acre, payments per acre are not correlated with other, unobserved factors affecting concentration growth (εi).

The smooth functions were estimated using "loess," short for "local polynomial regression," which fits the smooth functions by estimating polynomial functions using points local to each fitted point, with local points weighted more heavily than further points. The smooth functions are estimated jointly with β using a Gauss-Seidel backfitting method, as described and implemented by Hastie. The software package used was the public domain package 'R' with the 'gam' package written by Hastie (www.r-project.org). See this reference for more details about the procedure.

The key modeling decision concerns the share of points considered local to each fitted point on the smooth functions. For the models, each point along the smooth functions was estimated using 5 percent of the ZIP Codes, which is the smallest share that was computationally feasible for the two-dimensional spatial surface using the hardware and software. Summaries of the models with payment quintiles are reported in appendix table A1, excluding the parametric components reported in table 7.

Appendix table A2 reports a series of alternative long-panel estimates that illustrate the general robustness of the link between payment levels and concentration growth. The first set of estimates in the table are ordinary least squares (OLS) estimates with the control functions $f()$, $g_c()$, $g_a()$, and $g_s()$ approximated using orthogonal polynomials of varying orders. The table reports the estimated payment-group effects conditional on the polynomial controls. The orders of single-variable polynomials ($g_c()$, $g_a()$, and $g_s()$) are the given in the first column; the orders of the two-variable spatial function $f()$ are twice the number in the first column. Results can be compared to the unadjusted estimates (without controls), replicated from table 6 at the top of the table.

The second set of estimates in table A2 are robust regressions, or M estimates, fitted using iterated re-weighted least squares (IWLS). The weights are derived from penalty functions that reduce or eliminate the influence of outliers. The estimates are derived using Tukey's biweight proposal. The estimates were implemented using the public domain software "R" and the "MASS" package written by Venables and Ripley. More details can be found in Venables and Ripley and in the R documentation for the "rlm" function in the MASS package. Estimates were derived using the default method for the "MM" option. Robust methods are more computationally expensive than OLS, which limits the order of polynomials to 10 for both the single variable and two-variable spatial controls.

Table A1

Summary of generalized additive model estimates

Panel	Factor	Non-parametric degrees of freedom	F-value	Goodness of fit
Cropland concentration 1987-1992	Spatial surface (x,y)	73.0	16.3	Adj. R^2 = 0.153 Est. VAR(ε) = 0.281
	Beginning-year concentration (c_0)	35.5	58.0	
	Ratio of cropland to ZIP Code area (a_0)	35.8	10.7	
	Crop sales per acre of cropland (s_0)	36.3	2.2	
Cropland concentration 1992-1997	Spatial surface (x,y)	72.9	13.4	Adj. R^2 = 0.130 Est. VAR(ε) = 0.294
	Beginning-year concentration (c_0)	35.2	65.0	
	Ratio of cropland to ZIP Code area (a_0)	35.7	11.4	
	Crop sales per acre of cropland (s_0)	36.1	4.5	
Cropland concentration 1997-2002	Spatial surface (x,y)	71.4	13.1	Adj. R^2 = 0.135 Est. VAR(ε) = 0.348
	Beginning-year concentration (c_0)	35.2	66.5	
	Ratio of cropland to ZIP Code area (a_0)	35.6	11.0	
	Crop sales per acre of cropland (s_0)	35.7	3.2	
Cropland concentration Long panel 1987-2002	Spatial surface (x,y)	88.9	23.6	Adj. R^2 = 0.212 Est. VAR(ε) = 0.49
	Beginning-year concentration (c_0)	45.5	42.9	
	Ratio of cropland to ZIP Code area (a_0)	45.5	13.3	
	Crop sales per acre of cropland (s_0)	46.0	4.7	

Notes: Estimates and standard errors for the parametric components of the models (the payment group factors) are reported in table 6.
All F-tests for nonparametrically estimated smooth functions are statistically significant with at least 0.1-percent confidence.

Table A2

Summary of alternative model estimates

	Predicted 1987-2002 percentage growth in cropland concentration by payment level						
	No Payments	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	R-squared
Unadjusted predictions (“Long panel” from table 6 and figure 6)							
	-1.6	17.9	20.8	32.0	42.7	49.6	
Adjusted predictions using generalized additive model (“Long panel” from table 7 and figure 7)							
	-8.3	20.6	26.3	34.9	41.7	46.2	*0.212*
Predictions adjusted using OLS polynomial regressions[1]							
Order = 1	3.9	18.8	21.3	33.0	44.0	48.4	*0.104*
Order = 2	6.2	20.1	21.1	32.2	42.5	47.3	*0.123*
Order = 3	6.0	19.8	21.2	31.9	42.4	48.2	*0.146*
Order = 4	2.1	19.6	22.3	32.8	43.0	49.7	*0.167*
Order = 5	-1.4	19.5	23.6	34.3	43.7	49.7	*0.177*
Order = 6	-2.8	21.0	25.2	34.7	42.7	48.6	*0.188*
Order = 7	-4.1	21.4	26.4	35.4	42.7	47.7	*0.200*
Order = 8	-4.4	22.0	27.2	35.8	42.3	46.5	*0.210*
Order = 9	-4.6	21.9	27.5	35.7	41.8	47.2	*0.221*
Order = 10	-5.2	22.6	28.2	35.9	41.2	46.8	*0.231*
Order = 11	-4.9	22.6	28.6	35.8	40.9	46.6	*0.237*
Order = 12	-4.6	22.6	28.4	35.6	41.1	46.5	*0.242*
Predictions adjusted using robust “MM” polynomial regressions[2]							Estimated VAR(ε)
Order = 1	-1.4	16.8	18.5	34.4	48.3	52.9	*0.403*
Order = 2	9.0	22.1	20.5	31.6	41.8	44.5	*0.389*
Order = 3	7.3	21.1	20.3	32.2	42.5	46.1	*0.387*
Order = 4	5.8	20.3	20.1	32.8	43.0	47.5	*0.384*
Order = 5	2.9	20.1	21.7	33.8	43.3	47.6	*0.379*
Order = 6	0.0	20.2	22.9	34.1	43.8	48.4	*0.369*
Order = 7	-2.9	20.2	23.9	34.5	44.4	49.5	*0.365*
Order = 8	-4.4	20.4	24.4	35.0	44.6	49.4	*0.361*
Order = 9	-5.7	20.9	25.0	34.9	44.3	50.0	*0.362*
Order = 10	-5.3	21.3	25.2	34.9	44.0	49.3	*0.359*

Notes: The table reports estimated effects of the payment quintiles on concentration growth after controlling for location and concentration, sales per acre of cropland, and the ratio of cropland to area in each ZIP Code in the beginning year of each panel. Each row of the table reports predictions from an alternative specification.

[1]OLS polynomial regression include polynomials of each control variable for the given order except location. Location is given by two coordinates determined via an Albers equal area projection of the ZIP Code centroids. A polynomial spatial surface of two times the given order is estimated for each regression. For example, for “Order=12,” a 12th-order, single-dimension polynomial is estimated for beginning-year concentration, sales per acre of cropland, and the ratio of cropland area to ZIP Code area; and a 24th-order two-dimensional polynomial is estimated over the two location coordinates.

[2]Polynomials for the robust regressions are limited to order 10 for all controls due to computational limitations and to ensure numerical stability. R-squared is not a well defined concept for robust regressions. Instead, we report the estimated variance of the uncorrupted error.