Supplementary Information for

Impact of Transnational Land Acquisitions on Local Food Security and Dietary Diversity

Marc F. Müller, Gopal Penny, Meredith T. Niles, Vincent Ricciardi, Davide Danilo Chiarelli, Kyle Frankel Davis, Jampel Dell'Angelo, Paolo D'Odorico, Lorenzo Rosa, Maria Cristina Rulli, Nathaniel D. Mueller

Corresponding Author: Marc F. Müller.
E-mail: mmuller1@nd.edu

This PDF file includes:

Figs. S1 to S9
Tables S1 to S7
References for SI reference citations

S1. Selection of Land Deals

The considered land deals were obtained by filtering the complete Land Matrix dataset (approximately 1900 deals) according to the following criteria:

- Selected deals are described as 'concluded', 'in startup phase' or 'in production' with transition into these status types after the year 2000.
- Selected deals are transnational, greater than 200 hectares and entail a transfer of user rights from smallholders and communities to commercial users.
- Selected deals involve the sale, lease, or concession of land. Deals focusing on contract farming only are excluded.
- The purpose of the deal reported in the Land Matrix includes one or more of the following key words: 'Food Crops', 'Agriculture Unspecified', 'Non-food agricultural commodity', 'Plantation' and 'Biofuel'.
- The spatial accuracy of the deal's location(s) is described as 'Coordinates' or 'Exact Location' in the Land Matrix dataset.

Using the above criteria reduces the 1900+ deals of the full land matrix dataset to an initial subset of 169 land deals in 207 distinct locations (single deals can concern several locations). Seven land deals were further removed because their approximate coverage (see following section) almost entirely overlaps with that of another considered deal (2 deals), or with that of an excluded (i.e. non-agricultural) deal from the Land Matrix (5 deals). One further deal was removed because the provided coordinates correspond to a major urban agglomeration. Another deal was removed because our land use analysis failed to detect any cropland within the approximated coverage of the deal in both 2005 and 2015 (hence we are not confident in the accuracy of our coverage estimate). One last deal was removed because its location country (South Sudan) did not exist at the beginning of the considered period, which presented challenges in terms of matching the diverse considered datasets. These steps allowed us to obtain our final sample of 160 land deals, which we considered in the targeting and crop type transition analyses. A further 29 deals were removed for the crop cover change analysis because their contracting date lies outside of the 2005-2015 period covered by the remote sensing crop cover data that we use. In contrast, the full set of 160 deals was considered for the other geospatial analyses (targeting and crop type transition), which rely on crop types estimated for circa 2000 (see Methods section in the main document). A full list of the original 169 deals extracted from the land matrix, along with their exclusion rationale (when applicable) is openly available in the data repository https://doi.org/10.7274/r0-ycpf-qh53.

The filtering criterion on spatial accuracy is the most restrictive and rules out about 85% of the approximately 1900 deals of the full Land Matrix dataset. Having a high spatial accuracy is a critical requirement for our analysis, which relies on geo-spatial queries of a wide range of geo-referenced datasets. However, discarding 85% of the available land deals comes at the expense of a reduced variety in the selected sample of deals and a lower statistical power. More importantly, potential biases may emerge if the availability of spatial data is correlated with the food-security impact of the associated deals. For instance, our study will underestimate the average impact of land deals on food security if the particular deals that have the highest impact also tend to not have spatial data available. Unfortunately, the correlation between the availability of spatial information and the deals' impact on food security is impossible to measure because our approach to estimate the food security impacts of land deals precisely relies on the availability of spatial information about each deal. Hence, while we have no reason to believe that such biases did emerge, we are not able to formally rule them out.

S2. Deal Extents and Control Regions

The spatial extent of each deal was approximated as a disk with area and center coordinates equivalent to the corresponding values reported in the Land Matrix. Note that these disks are just an approximation of the actual spatial extents of the deals, which are not reported in the Land Matrix (see following Section for uncertainty assessment). Deals associated with more than one location were split into multiple disks of equal areas that sum to the total area of the deal. Spatial subtractions were carried out to remove potential overlaps with disks corresponding to Land Matrix deals that were not included in the analysis (e.g., because they are not focused on agriculture). Control regions were defined as donut-shaped regions at increasing distances from the deal centroid. Overlap with disks corresponding to other deals from our sample, or with donuts corresponding to control regions of another type (e.g., overlaps between B1 and B3) were spatially subtracted from the control regions. Consequently, control regions could not be defined for some of the deal-disks, where the donut-shaped control regions are entirely contained within the control region or disk of another deal. These deals were removed to form our final sample of 160 Land Matrix deals. A small number of deals were in the vicinity of other deals and the deal disks overlapped. These deals did not receive special treatment, so that meta-information regarding the deals were calculated and analyzed independently, meaning some pixels (approximately 1.1% of all pixels within deal disks) were counted more than once.

S3. Location Uncertainty

A validation sample of 222 Cambodian land deals with known spatial extents was obtained from the Economic Land Concessions (ELC) dataset (https://data.opendevelopmentcambodia.net/), and used to evaluate the locational uncertainties of the analysis.

We first evaluated the uncertainty associated with our approximation of the extent of deals and control regions as disks and tors (donuts), respectively. To do so, we applied the same approximation to the validation sample of 222 Cambodian deals

(Figure S1 A) and evaluated the extent to which the actual deals overlap with approximated deals and control regions. We found that 75% of the actual surface area of deals overlap with the approximated disks. (Equivalently, since approximated disks are constructed so as to have the same surface area as actual deals, 75% of the surface area of approximated disks overlap with the actual surface area of deals.) We also found that 21%, 0.06% and 0% of the surface area of control regions B1, B2 and B3 (respectively) overlap with real deals. This suggests that there might be concern for contamination of control region B1, which would explain some of the identified 'spillover' effect in the region, if applicable. However, this concern vanishes when considering control regions B2 and B3. Regarding the spatial approximation of the deal regions themselves, the 75% overlap raises confidence that much of the effect of the deal on land cover is captured by our analysis. If anything, the estimated effect is conservative, owing to the fact that 25% of the deal-disk approximation may covers non-deal region.

Second, we evaluated the uncertainty associated with the fact that the sample of deals selected from the Land Matrix dataset is likely not exhaustive. The risk is that the control regions in the analysis may (partly) cover land that is covered by deals that are not included in the analysis. We use the sample of 222 Cambodian deal, most of which are not included in our Land Matrix sample, to investigate this effect. We find that out-of-sample land deals cover 1.5%, 0.1% and 0.1% of the surface area of control regions B1, B2, B3 that are associated with the 21 Cambodian deals of our Land Matrix sample. This suggests that out-of-sample contamination is of little concern to the analysis. Note that, even if out-of-sample contamination were a concern, it would make our estimated effect of LSLA deals on crop cover change conservative.

The above analyses alleviate the concern that the control area B3 might be directly contaminated by containing part of the actual extent of a deal (whether the deal associated with the control area itself, or an out-of-sample deal). However, it is possible that crops in B3 are affected by spillovers from the corresponding deal. This possibility specifically applies to the crop expansion analysis, where spillovers of deals to B1 and B2 have been detected (Figure 2B of the main document). To alleviate this concern, we added the administrative region that contains each deal as an additional control area. We used the Global Administrative Areas (GADM) dataset (https://gadm.org/) and selected the Level 3 (or Level 2, if Level 3 was unavailable) administrative area containing the centroid of each deal as a fourth control area, in addition to B1, B2 and B3.

S4. Visual Categorization

LSLA disks were imported into Google Earth Pro. For the most recent image available for an individual LSLA, we recorded the year of the image and the presence or absence of irrigation pivots or the development of large-scale fields with regular polygonal dimensions. We repeated this analysis for the available image with year closest to year 2000. The visual categorization was carried out independently three times by a different analyst, and the outputs were cross-checked. At the start of the study period, most LSLAs contained mosaic landscapes typical of smallholder fields, whereas 7% of the deals showed evidence of established commercial agricultural enterprise. This proportion increased to 42% for the most recent available images. Signs of pivot irrigation on the most recent images were observed for 8% of the deals, most often on newly established commercial farms.

S5. Crop cover data and resampling uncertainty

The two global crop cover datasets used in the analysis are based on different grids with different resolutions. The 2005 Global Land Cover Product (Globcover2005) was obtained from http://due.esrin.esa.int/page_globcover.php via Google Earth Engine. It has a resolution of 300 meters with four crop intensity categories indicating the percentage of each pixel covered by crops in 2005 ($< 20\%$, $20-50\%$, $50-70\%$ and $> 70\%$). The 2015 Global Food Security-support Analysis Data at 30 meter resolution (GFSAD30) for South Asia, Southeast Asia, Africa, Europe and South America were merged to obtain a consistent 30-meter resolution grid with binary pixels indicating cultivation status in 2015 (crop vs non-crop). We combined the 2005 and 2015 datasets as follows.

The 30-meter resolution binary (crop, non-crop) GFSAD30 data grid was aggregated into 300-meter pixels that match the Globcover2005 grid. Actual crop cover fractions were computed for each 300-meter grid cell, which was then classified as one of the four Globcover2005 categories ($< 20\%$, $20-50\%$, $50-70\%$ and $> 70\%$ crop cover). The two datasets in Globcover2005 format (2005 and 2015) were then used for two purposes. First, to investigate changes in spatial concentration, as the number of 300-meter pixels with at least 70% crop coverage (Figure 2C in the main document). Second, to select permanently cropped pixels (i.e. pixels with a crop coverage of 70% or more in both 2005 and 2015) for the NDVI analysis described in the following section.

Reversely, the 300-meter resolution categorical Globcover2005 dataset was converted into a 300-meter continuous-scale crop-cover fraction grid. To do so, we multiply each cell by a coefficient corresponding to the average crop cover fraction of its category, estimated using actual crop cover fractions from the aggregated 2015 data (see previous paragraph). Estimated crop cover fractions were used to estimate the effect of LSLA deals on cropland expansion. In the main analysis (Figure 2B in the main document), multiplying coefficients were obtained by taking the average crop cover fractions by Globcover2005 category for each region (Asia, Africa, Europe and South America). However, we replicated the expansion analysis using a *global* average crop fraction with very similar results (Figure S2, bottom).

To further assess the uncertainty associated with the resampling procedure, we plotted (i) the *estimated* 2015 crop fraction by deal obtained by aggregating the 30-meter GFSAD30 data into 300-meter Globcover2005 categories and then multiplying by the relevant crop cover averages, against (ii) the *actual* crop fraction computed directly from 30-meter GFSAD30 data. Results, displayed in Figure S2 (top-left), align along the 1:1 line suggesting that the associated resampling uncertainty is negligible. We constructed another similar scatterplot, this time letting the multiplying coefficient vary by region, with results pointing to a similar conclusion (Figure S2, top-right).

S6. Crop Intensity Analysis based on the Normalized Difference Vegetation Index

To further investigate the hypothesized effect of land deals on agricultural intensification, we use remotely-sensed Normalized Difference Vegetation Index (NDVI) as a proxy of biomass production. We focus on the subset of 300-meter pixels that were classified as $> 70\%$ cropped in both 2005 and 2015. Focusing on persistently cropped pixels allows to (partially) control for the effect of land use changes on the area-averaged NDVI. For each pixel and each year, we calculated the mean and maximum NDVI value captured by the space-borne MODIS sensor (MOD13A1.005 Vegetation Indices 16-Day L3 Global 500m obtained from NASA's Land Processes Distributed Active Archive Center via Google Earth Engine). We then took the spatial mean of the obtained annual NDVI value for each deal and control area, so as to generate annual time series of (i) spatially averaged mean annual NDVI and (ii) spatially averaged maximum annual NDVI for persistently cropped pixels.

The NDVI analysis focused on a subset of 61 deals that were selected from our original sample so as to contain at least one 300-meter pixel that satisfies all of the following conditions:

1. It should be 'permanently cropped', here defined as being classified in the $> 70\%$ cropped category in both 2005 and 2015.

2. It should contain a minimum of 3 'acceptable' annual NDVI estimates (mean or max), here defined as having more than 50% of that year's observations with no quality flag in the original MODIS dataset.

3. It should contain at least one NDVI estimate taken before and one NDVI estimate taken after the corresponding deal was signed.

We used the NDVI time series in the deal and control regions to perform separate non-parametric regressions (loess) against time before and after the contract year. Similar to regression discontinuity (e.g., 1), this approach allows the temporal dimension of the effect to be visualized. A discontinuity in either level or temporal trend a the deal year that does not appear in the control region are suggestive of a causal effect of deals on NDVI values. Results in Figure S4 show that, although NDVI fluctuates (e.g., due to climate variability) and may display slight increasing trends (e.g., max annual NDVI in Europe and South America), no discontinuity arises at or around the deal year. Thus, the analysis shows no evidence of a relationship between land deals and biomass production in persistently cropped regions.

S7. Household Survey Analysis: Estimation details and Robustness Checks

We conducted the following robustness checks on the household survey analysis.

- **Misallocation of households to deals:** It is possible that the control zones B1-B3 associated with different land deals overlap spatially (see "Deal Extents and Control Regions" above). In our analysis, this happens for 6 (three pairs) of the 28 land deals for which DHS household data are available. We address these conflicts by randomly assigning the corresponding households to either of the two overlapping land deals. To alleviate potential concerns that this arbitrary (albeit random) assignment might drive our results, we reproduce our preferred specifications ((1) and (2) on Table S6) 1000 times with different random land deal assignments for households in overlapping deal regions. For specification (1), deal assignments are used as a clustering variable in our original analysis (as opposed to a regression covariate). Consequently, only the standard errors (not the regression estimates themselves) vary across repetitions. We find that the regression coefficient for *afterLSLA* remains statistically significant ($p < 0.01$) for all 1000 repetitions. In specification (2), estimates for the *Year:afterLSLA* coefficient lie between -0.28 and -0.52 for all repetitions (Figure S9, panel C), to be compared to an estimate of -0.353 in our main analysis (Table S6). Again, all estimates are statistically significant at the same confidence level ($p > 0.1$) as the original analysis.

- **Alternative Dietary Diversity Scores:** The dietary diversity score used in the main analysis (Table S6) is prescribed by the Food and Agriculture Organization and based on intake from 10 particular food groups (2), namely: (i) cereal grains, (ii) white tubers and root foods, (iii) dark leafy greens, (iv) vitamin A rich vegetables/tubers, (v) vitamin A rich fruits, (vi) other fruits and vegetables, (vii) meat and fish foods, (viii) eggs, (ix) legumes/nuts/seeds, and (x) milk and milk products. We test the robustness of our results to the considered set of food groups by considering an alternative dietary diversity metric, the Minimum Dietary Diversity Score (MDDS), prescribed by the World Health Organization (3) and based on a (different) set of 7 food groups: (i) grains, roots, and tubers, (ii) vitamin A - rich fruits and vegetables, (iii) other fruits and vegetables, (iv) flesh foods (meat, organs, fish), (v) eggs and legume, (vi) nuts and seeds, and (vii) dairy. Regression results are shown in Table S7 and are comparable to the main specifications. The smaller amplitude of the estimated effect is likely attributable to the smaller amplitude of the dietary diversity score (7 points instead of 10 points). Most importantly, all coefficients remain statistically significant at a confidence level that is identical to the corresponding original specifications in Table S6.

- **Error in Tree Cover observations:** Tree cover, which has a highly significant relation to dietary diversity in our regressions on Table S6, may be challenging to accurately estimate in sparsely forested West Africa (4), where most of the considered deals are located. We assess the sensitivity of our results to detection errors in tree cover by adding to our tree cover observations randomly generated synthetic errors that are drawn from independent uniform distributions. Results shown in Figure S9 (panels A and B) suggest that our main results (the coefficients for *afterLSLA* and *Year:afterLSLA* in regressions (1) and (2) on Table S6) are little sensitive to the amplitude of the added errors, represented as the bounds of the uniform distribution from which they are drawn. All estimates lie within 1% of the original regression coefficient and exhibit an identical statistical significance (namely $p < 0.01$ and $p < 0.1$, respectively).

- **Spatial displacement of DHS Clusters:** Prior to their release, rural DHS clusters are displaced by a random distance of 0–5 km in any direction, with $\sim 1\%$ of coordinates randomly shifted by up to 10 km. Because of this, DHS recommends that environmental data be averaged over a 5–10 km buffer around each DHS rural location. In our analysis, we account for cluster displacement as follows for the various covariates that we consider. The livestock density data that we use has a native resolution of approximately 5 km (see Table S4), which matches the 5-10 km averaging buffer suggested by DHS, so no further averaging is applied. Population density is averaged within a radius of 5 km around each cluster coordinate, as recommended by DHS. Tree cover is averaged over a 20km radius around the DHS cluster, as an approximation of the distance that one can travel on foot during one day to forage or hunt. Lastly, land deals are associated with all DHS households falling within a distance of four times their disk radius (i.e. zones 'Deal', 'B1', 'B2' and 'B3'). The smallest land deal of our sample that has DHS households associated with it has a radius of 1.3 km (Land Matrix ID 4060), in which case all DHS clusters within a distance of $1.3 \times 4 = 5.2$ km are selected. This ensures that all DHS clusters located within the area covered by land deals are included in the analysis, in spite of their random displacement.

References

1. Jaynise M Pérez Valentín and Marc F Müller. Impact of hurricane maria on beach erosion in puerto rico: Remote sensing and causal inference. *Geophysical Research Letters*, 47(6), 2020.
2. Gina Kennedy, Terri Ballard, and Marie Claude Dop. *Guidelines for measuring household and individual dietary diversity*. Food and Agriculture Organization of the United Nations, 2011.
3. World Health Organization et al. Indicators for assessing infant and young child feeding practices: part 2: measurement. 2010.
4. Martin Brandt, Compton J Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, et al. An unexpectedly large count of trees in the west african sahara and sahel. *Nature*, pages 1–5, 2020.
5. Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, SV Stehman, Scott J Goetz, Thomas R Loveland, et al. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013.
6. Günther Fischer, Freddy O Nachtergaele, Sylvia Prieler, Edmar Teixeira, Géza Tóth, Harrij Van Velthuizen, Luc Verelst, and David Wiberg. Global agro-ecological zones (gaez v3. 0)-model documentation. 2012.
7. Walter V Reid, Harold A Mooney, Angela Cropper, Doris Capistrano, Stephen R Carpenter, Kanchan Chopra, Partha Dasgupta, Thomas Dietz, Anantha Kumar Duraiappah, Rashid Hassan, et al. *Ecosystems and human well-being-Synthesis: A report of the Millennium Ecosystem Assessment*. Island Press, 2005.
8. Center for International Earth Science Information Network CIESIN Columbia University and Centro Internacional de Agricultura Tropical CIAT. Gridded population of the world, version 3 (gpwv3): Population density grid. 20200814 2005. URL https://doi.org/10.7927/H4XK8CG2.
9. Center for International Earth Science Information Network CIESIN Columbia University and Information Technology Outreach Services ITOS University of Georgia. Global roads open access data set, version 1 (groadsv1). 20200814 2013. URL https://doi.org/10.7927/H4VD6WCT.
10. WHO and UNICEF. Global water supply and sanitation assessment 2000 report. *World Health Organization and United Nations Children's Fund (WHO and UNICEF)*, 2000.

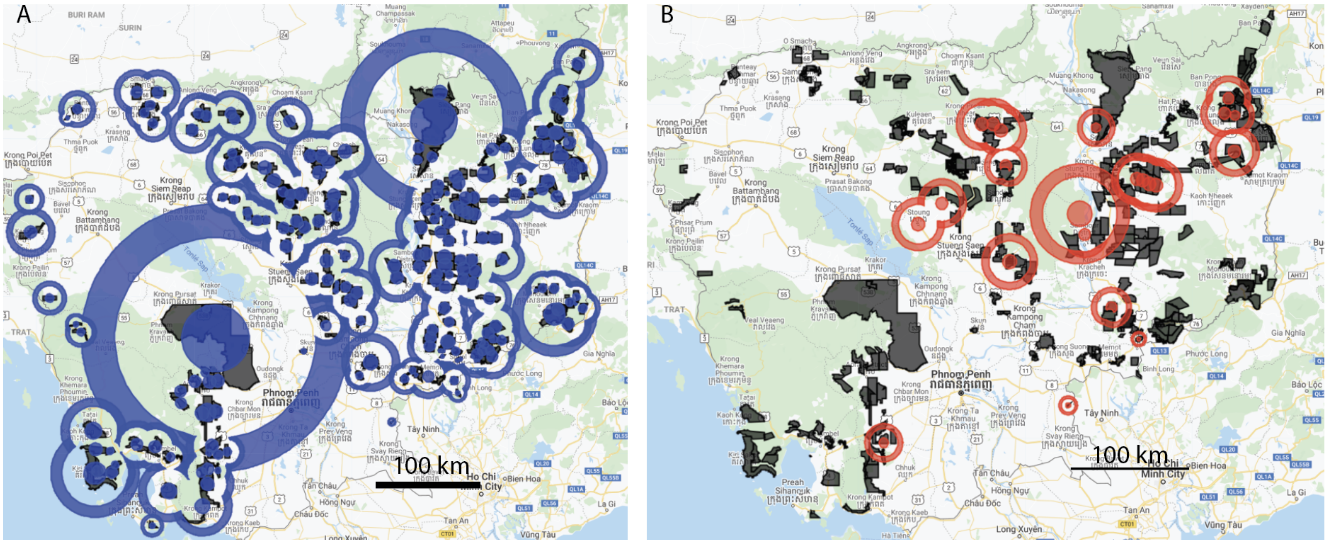

Fig. S1. ***Locational Uncertainties.*** A. Locational approximation of deals (disks) and control regions B3 (tors) applied to a set of 222 deals in Cambodia with known extents. Approximated deals (disks) overlap with 75% of the coverage of real deals and real deals overlap with 75% of the coverage of approximated deals. Note that these two statements are equivalent because approximated deals are constructed to have the same surface area of real deals. Control areas B1, B2 and B3 overlap with real deals by 21%, 0.06% and 0%, respectively. B. Overlap between selected Land Matrix deals (red disks), and associated control regions (here B3, red tors), and the 222 Cambodian deals with known location (black). Overlap between LandMatrix control regions and out-of-sample Cambodian deal was is 1.5%, 0.1% and 0.1% of the control region's surface area for B1, B2 and B3, respectively.

Table S1. Surface area (in millions of hectares) of crop cover land in 2005 and 2015, and total land area by region and area type. Admin. Area designates the level 2 or level 3 (if applicable) administrative area that contains the land deal centroids.

	Deal			B1			B2			B3			Admin. Area		
	$Crop_{05}$	$Crop_{15}$	Tot.	$Crop_{05}$	$Crop_{15}$	Tot.	$Crop_{05}$	$Crop_{15}$	Tot.	$Crop_{05}$	$Crop_{15}$	Tot.	$Crop_{05}$	$Crop_{15}$	Tot.
Africa	0.19	0.22	2.40	0.79	0.75	10.63	1.28	1.13	16.74	1.64	1.35	20.33	4.97	4.64	98.80
Asia	0.11	0.16	0.58	0.46	0.56	1.85	0.77	0.85	2.83	1.07	1.18	3.82	3.30	2.13	10.62
Lat. Am.	0.24	0.28	0.55	0.65	0.88	1.92	0.77	1.26	3.19	0.74	1.45	4.45	1.68	2.42	6.80
Europe	0.30	0.39	0.54	1.33	1.78	2.21	2.06	2.75	3.50	2.60	3.44	4.60	3.86	5.01	7.41

Table S2. Crop cover fraction in 2005 and 2015 (CCF_{2005}, CCF_{2015}) for the 6 largest deals of the sample (top 4 percentile). Most deals are in Sub-Saharan Africa and exhibit a *decrease* in crop cover fraction (in bold).

LandMatrix ID	Country	area of deal [ha]	CCF_{2005}	CCF_{2015}
1166	Congo, Rep.	470000.00	0.04	**0.01**
3393	Ghana	400000.00	0.04	**0.02**
746	Brazil	300000.00	0.61	0.66
1037	Ukraine	242400.00	0.58	0.79
1388	Liberia	220000.00	0.19	**0.01**
5586	Cameroon	195921.00	0.02	**0.01**

Table S3. Number of households per country and DHS survey year. The number of land deals that overlap spatially with surveyed households, the number of households surveyed before and after the implementation of the corresponding deal are also given

	2006	2007	2008	2009	2010	2011	2012	2013	2014	Deals	HH Pre-deal	HH Post-deal
Cameroon	0	0	0	0	0	22	0	0	0	1	0	22
Ethiopia	0	0	0	0	0	7	0	0	0	1	0	7
Ghana	0	0	508	0	0	0	0	0	0	7	502	6
Liberia	0	1,402	0	0	0	0	0	1,269	0	2	1,402	1,269
Malawi	0	0	0	0	121	0	0	0	0	1	0	121
Mali	0	0	0	0	0	0	32	0	0	1	0	32
Senegal	0	0	0	0	23	65	0	0	0	2	0	88
Sierra Leone	0	0	0	0	0	0	0	746	0	5	0	746
Swaziland	201	5	0	0	0	0	0	0	0	1	206	0
Uganda	21	0	0	0	0	8	0	0	0	2	0	29
Zambia	0	18	0	0	0	0	0	64	8	5	18	72
Total	222	1,425	508	0	144	102	32	2,079	8	28	2,128	2,392

Variable Name	Variable Description	Unit	Source
Diet diversity	Mean Individual Diet Diversity Score (IDDS) of children per household	IDDS 0-10 (see caption)	DHS
Tree cover	Percent tree cover within a 20km radius of cluster	% tree cover, ca. 2000	Hansen et al. 2013 (5)
Livestock density	Ruminant livestock density, circa 2000	TLUS 1-9 (see caption)	FAO (6)
Distance to urban center	Distance to its nearest settlement with 5000+ inhabitants	meters, ca. 2000	Millenium Assess. (7)
Distance to road	Distance to nearest road	meters, ca. 2000	NASA SEDAC (8)
Population density	Average population density within 5km buffer around cluster	people per km^2 (see caption)	NASA SEDAC (9)
Time to water	Time to get to water source	minutes, DHS survey year	DHS
Child age	Mean Age of children below 5 years per household	month, DHS survey year	DHS
Education of household head	Education years of head of household	years, DHS survey year	DHS
Improved toilet	WHO definition of improved sanitation	Binary, DHS survey year	DHS
Improved water	WHO definition of improved water supply	Binary, DHS survey year	DHS
Male household head	Household head male	Binary	DHS
Poorest household	within the community's poorest wealth quintile	binary, DHS survey year	DHS
Richest household	within the community's richest wealth quintile	binary, DHS survey year	DHS

Table S4. Diet diversity scores are computed for the Demographic and Health Surveys (DHS) year based on intake of 10 foods including cereal grains, white tubers and root foods, dark leafy greens, vitamin A rich vegetables/tubers, vitamin A rich fruits, other fruits and vegetables, meat and fish foods, eggs, legumes/nuts/seeds, and milk and milk products. Tree cover is obtained from the Hansen dataset (5), which provides tree cover estimates (as a relative value between 0 and 100%) per 30-m pixel in the year 2000. The tree cover value considered in our analysis is obtained by simply taking the spatial average of this value, within a 20km radius disc around each DHS household coordinate. Tropical livestock unit (TLU) scores, circa 2000: 1 if less than 1 TLU; 2 if between 1 and 5 TLU; 3 if between 5 and 10 TLU; 4 if between 10 and 20 TLU; 5 if between 20 and 50 TLU; 6 if between 50 and 100 TLU; 7 if between 100 to 200 TLU; 8 if more than 200 TLU; 9 if water cover. Population density was taken in 2000, 2005 and 2010 if the household survey was taken between 2000 and 2004, 2005 and 2009, and 2010 and 2013, respectively. *Improved Toilet* and *Improved Water* refer to the indicators defined by WHO/UNICEF Joint Monitoring Programme to evaluate Target 7C of the millennium development goals (10). Therein, improved water supply refers to access to the following sources of water for drinking: household water connection, a public standpipe, a borehole, a protected dug well or spring, or a rainwater collection system. The term excludes unprotected wells, unprotected springs, vendor-provided water, bottled water or tanker truck-provided water. Improved sanitation refers to connection to a public sewer or a septic system, access to a pour-flush latrine or to a simple pit latrine or a ventilated improved pit latrine. It excludes service or bucket latrines (where excreta are manually removed), public latrines or latrines with an open pit. In our analysis, the corresponding variables are coded as "1" if the household has access to improved water (respectively sanitation) and "0" otherwise.

Table S5. Descriptive statistics of DHS household characteristics

Household Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Number of Children	4,520	1.735	0.812	1	1	2	8
Tree Cover	4,520	18.330	15.739	0	0.4	29.9	80
Livestock density	4,520	15.367	19.608	0	5	20	360
Time to Water	4,520	2.568	1.590	1	1	3	9
Dist. to Urban center	4,520	19,911.680	21,830.280	0.000	895.531	30,302.840	89,095.220
Dist. to Road	4,520	1,176.025	2,167.525	0.820	170.100	1,157.470	20,084.210
Pop. Density	4,520	690.873	1,420.451	4.210	38.397	167.870	6,385.427
HH-Average Age of Children in Months	4,520	21.962	10.200	0	15	29.5	48
Edu. of HH Head	4,520	5.311	6.297	0	0	9	98
Improved Toilet	4,520	0.091	0.288	0	0	0	1
Improved Water	4,520	0.494	0.500	0	0	1	1
Sex of HH Head	4,520	0.688	0.463	0	0	1	1
Poorest HH	4,520	0.236	0.425	0	0	0	1
Richest HH	4,520	0.128	0.334	0	0	0	1

Table S6. Regression Results. Specifications (1) and (4) represent Ordinary Least Squares regression results for the 2,671 sampled DHS households in Liberia. They show that the mean (1) and minimum (4) individual dietary diversity score (IDDS) per household for children under 5 (*afterLSLA* coefficient) decreases by about a point, when comparing before and after the contract year of the corresponding land deal. The effect is significantly different from zero at the 99% confidence level for both specifications, with standard errors clustered by land deal. Specifications (2), (3) and (5) represent Reduced Maximum Likelihood estimates of a mixed effects linear regression of our full sample of 4,520 DHS households across 11 African countries. They show that time trends on the mean ((2) and (3)) and minimum (5) IDDS per household decreases by about half a point per year (*Year:afterLSLA* coefficient) when comparing the situation before and after the corresponding land deals. This effect is statistically significant, respectively at the 90% (2) and 99% ((3) and (5)) confidence levels, for all specifications. The mixed effect model controls for random effects by land deal. For all specifications, standard errors are given in parentheses and asterisks are used to denote statistical significance.

	Dependent variable:				
	Mean IDDS (FAO) per Household			Min IDDS (FAO) per Household	
	OLS	*linear mixed effects*		*OLS*	*linear mixed effects*
	Liberia	Full Sample	Full Sample	Liberia	Full Sample
	(1)	(2)	(3)	(4)	(5)
Year		0.171	0.165		0.214
		(0.190)	(0.180)		(0.179)
afterLSLA	−1.186***	707.660*	900.344**	−1.145**	784.667**
	(0.410)	(404.992)	(394.007)	(0.453)	(381.945)
Year:afterLSLA		−0.353*	−0.448**		−0.391**
		(0.202)	(0.196)		(0.190)
Year:afterLSLA:B3			0.322*		
			(0.192)		
Tree Cover	0.007	0.005**	0.003	0.007	0.006**
	(0.007)	(0.003)	(0.003)	(0.006)	(0.003)
Dist. to Urban center	0.00001**	0.00001***	0.00001***	0.00001**	0.00001***
	(0.00001)	(0.00000)	(0.00000)	(0.00001)	(0.00000)
Dist. to Road	0.0001***	0.00002	0.00001	0.0001***	0.00002
	(0.00001)	(0.00002)	(0.00002)	(0.00001)	(0.00002)
Pop. Density	0.0002***	0.0001***	0.0001***	0.0002***	0.0001***
	(0.00002)	(0.00003)	(0.00003)	(0.00002)	(0.00003)
Time to water	0.001	0.0003	0.00003	0.001	0.0004
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Livestock density	0.075	0.030	0.027	0.044	0.024
	(0.133)	(0.029)	(0.029)	(0.098)	(0.029)
Mean Age of Children	0.061***	0.062***	0.062***	0.062***	0.063***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Education years of HH head	0.018	0.017***	0.017***	0.017	0.018***
	(0.013)	(0.005)	(0.005)	(0.014)	(0.005)
Improved Toilet	0.009	−0.076	−0.106	−0.002	−0.056
	(0.046)	(0.131)	(0.131)	(0.050)	(0.133)
Improved Water	0.169*	0.215***	0.212***	0.196	0.237***
	(0.094)	(0.080)	(0.079)	(0.138)	(0.080)
Sex of HH head	−0.047	−0.057	−0.074	−0.011	−0.040
	(0.061)	(0.070)	(0.070)	(0.070)	(0.071)
Poorest Quintile	0.037	−0.047	−0.063	0.087	0.0005
	(0.103)	(0.083)	(0.083)	(0.104)	(0.084)
Richest Quintile	0.547***	0.556***	0.542***	0.581***	0.581***
	(0.044)	(0.118)	(0.118)	(0.066)	(0.120)
Year:B3			−0.188		
			(0.171)		
afterLSLA:B3			−646.704*		
			(385.097)		
B3			377.952		
			(342.324)		
Intercept	0.592	−341.676	−331.136	0.465	−429.043
	(0.391)	(381.959)	(360.603)	(0.309)	(359.522)
Observations	2,671	4,520	4,520	2,671	4,520

Note: *p<0.1; **p<0.05; ***p<0.01
Mixed Effects:Hierarchical random effects on intercept by deal and country
OLS: Standard Error clustered by land deal

Table S7. Replicated Regression Results with alternative diet diversity score, namely the minimum dietary diversity score (MDDS). Specifications (1) and (4) represent Ordinary Least Squares regression results for the 2,671 sampled DHS households in Liberia. They show that the mean (1) and minimum (4) MDDS per household for children under 5 (*afterLSLA* coefficient) decreases by about a point, when comparing before and after the contract year of the corresponding land deal. The effect is significantly different from zero at the 99% confidence level for both specifications, with standard errors clustered by land deal. Specifications (2), (3) and (5) represent Reduced Maximum Likelihood estimates of a mixed effects linear regression of our full sample of 4,520 DHS households across 11 African countries. They show that time trends on the mean ((2) and (3)) and minimum (5) MDDS per household decreases by about half a point per year (*Year:afterLSLA* coefficient) when comparing the situation before and after the corresponding land deals. This effect is statistically significant, respectively at the 90% (2) and 99% ((3) and (5)) confidence levels, for all specifications. The mixed effect model controls for random effects by land deals. For all specifications, standard errors are given in parentheses and asterisks are used to denote statistical significance.

	Dependent variable:				
	Mean MDDS per Household			Min MDDS per Household	
	OLS	*linear mixed effects*		*OLS*	*linear mixed effects*
	Liberia	Full Sample	Full Sample	Liberia	Full Sample
	(1)	(2)	(3)	(4)	(5)
Year		0.169	0.212		0.210
		(0.175)	(0.165)		(0.171)
afterLSLA	−0.829***	626.609*	838.855**	−0.804***	704.383*
	(0.207)	(370.562)	(358.553)	(0.244)	(362.190)
Year:afterLSLA		−0.312*	−0.418**		−0.351*
		(0.185)	(0.179)		(0.180)
Year:afterLSLA:B3			0.336**		
			(0.150)		
Tree Cover	0.009	0.005***	0.004**	0.008	0.006***
	(0.006)	(0.002)	(0.002)	(0.006)	(0.002)
Dist. to Urban center	0.00001*	0.00000***	0.00000***	0.00001*	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Dist. to Road	0.00004***	0.00001	0.00000	0.00004***	0.00001
	(0.00000)	(0.00001)	(0.00001)	(0.00000)	(0.00001)
Pop. Density	0.0002***	0.0001***	0.0001***	0.0002***	0.0001***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Time to water	0.001	0.0004	0.0003	0.001	0.0005
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Livestock density	0.021	0.022	0.024	0.002	0.016
	(0.089)	(0.023)	(0.023)	(0.065)	(0.024)
Mean Age of Children	0.047***	0.049***	0.049***	0.048***	0.049***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Education years of HH head	0.015*	0.016***	0.015***	0.014	0.016***
	(0.008)	(0.004)	(0.004)	(0.009)	(0.004)
Improved Toilet	0.034	−0.056	−0.074	0.027	−0.037
	(0.045)	(0.102)	(0.102)	(0.047)	(0.104)
Improved Water	0.156***	0.172***	0.172***	0.176**	0.189***
	(0.048)	(0.062)	(0.062)	(0.078)	(0.063)
Sex of HH head	−0.022	−0.048	−0.060	0.008	−0.034
	(0.018)	(0.054)	(0.054)	(0.022)	(0.055)
Poorest Quintile	−0.036	−0.065	−0.076	0.011	−0.026
	(0.111)	(0.064)	(0.064)	(0.111)	(0.065)
Richest Quintile	0.406***	0.468***	0.457***	0.431***	0.487***
	(0.024)	(0.092)	(0.092)	(0.033)	(0.093)
Year:B3			−0.249*		
			(0.132)		
afterLSLA:B3			−674.177**		
			(300.333)		
B3			500.143*		
			(265.706)		
Constant	0.593**	−338.336	−424.168	0.485**	−420.305
	(0.286)	(350.800)	(330.474)	(0.220)	(342.606)
Observations	2,671	4,520	4,520	2,671	4,520
Note:					*p<0.1; **p<0.05; ***p<0.01

Mixed Effects:Hierarchical random effects on intercept by deal and country

OLS: Standard Error clustered by land deal

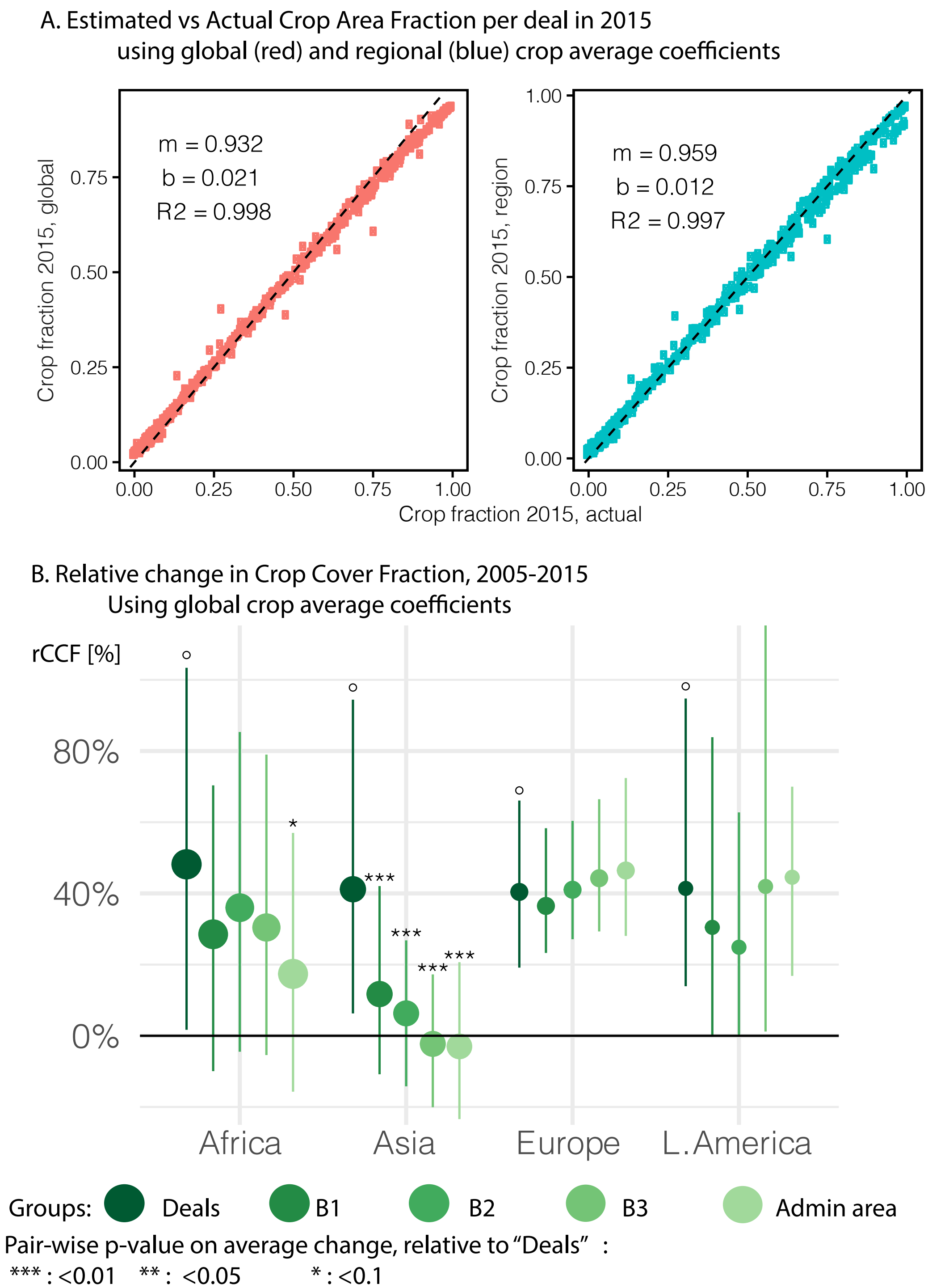

Fig. S2. *Effect of Resampling Uncertainties*: A. Comparison between approximated and actual crop fraction in 2015 for all land deals. Approximated crop fractions (y-axis) are obtained by (i) aggregating 30 meters binary (crop/non crop pixels) into 300m pixels with four crop categories ($<$20%, 20-50%, 50-70%, $>$70% crops) and (ii) multiplying the proportion of pixels of each category by a coefficient corresponding to the average crop cover of each category. Using a single set of globally averaged coefficient (red) or different sets of coefficients per continent (blue) does not increase the error between actual and approximated crop fractions in 2015. B. Similarly, using a single set of global coefficients to approximate crop fractions in 2015 does not substantially impact the results of the crop cover analysis in Figure 2B of the main document, where regional coefficients were used.

Fig. S3. ***Replication of the crop cover analyses with deals weighted by area***: Weighting the deals by area leads to smaller estimated outcomes in Africa and Latin America, both in terms of crop cover fraction (Panel A to be compared to Figure 2B in main document) and of spatially dense crop fraction (Panel B, to be compared to Figure 2C). However, removing the 9 (i.e. 5%) largest deal of the sample (Panels B and D) yields similar results to the unweighted analysis (Fig 2B and C). This suggests that area-weighted results are dominated by a small number of outliers: large deals, predominantly in Africa, where the crop cover fraction decreased between 2005 and 2015 (Table S2).

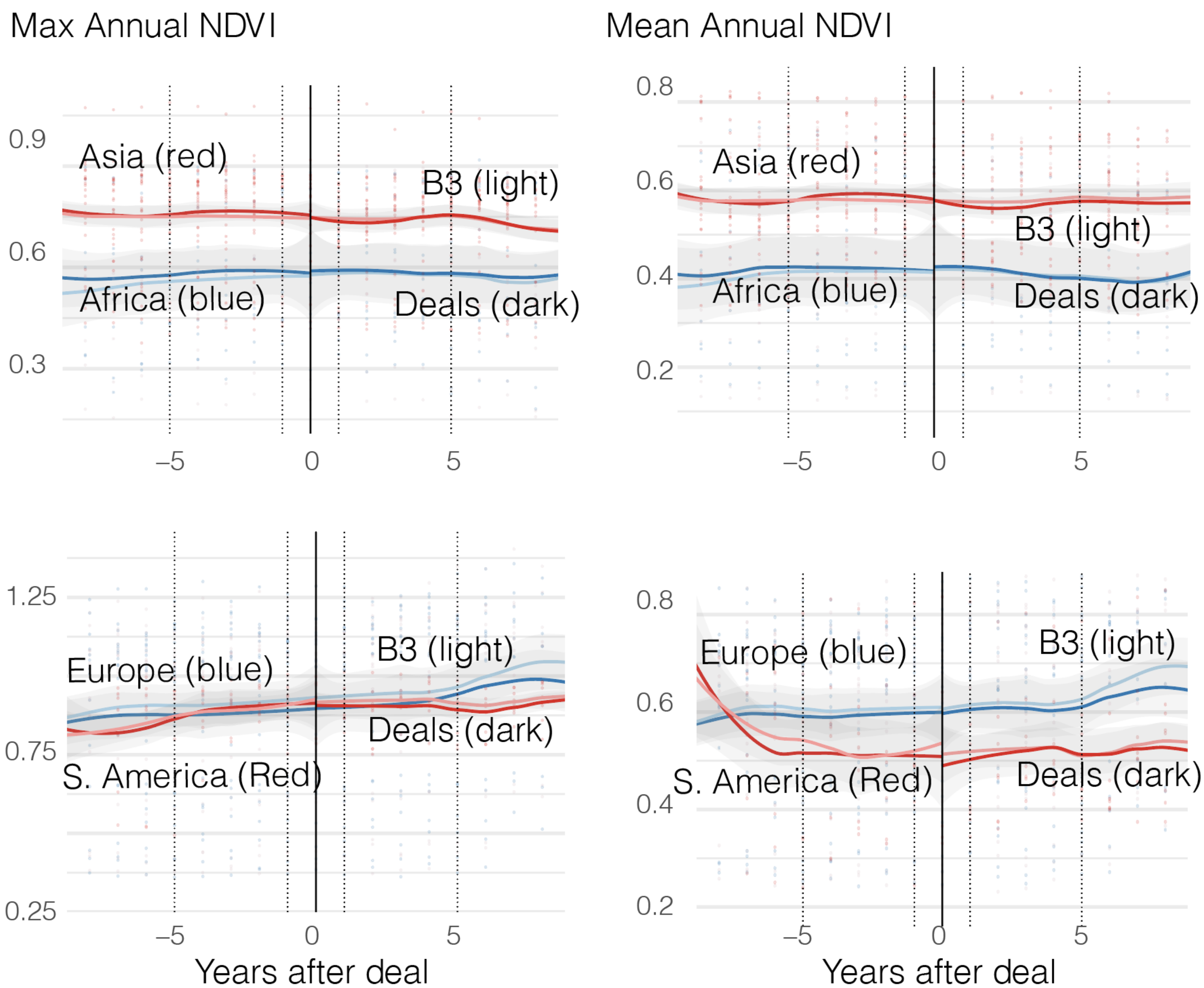

Fig. S4. ***Temporal discontinuity analysis***: Maximum (left) and mean (right) annual Normalized Difference Vegetation Index (NDVI) from MODIS satellite imagery, spatially averaged for each deal (dark) and control region B3 (light). NDVI estimates are plotted against the number of years before or after the contract date of each deal. Lines represent NDVI estimates from non-parametric regressions fitted independently to observations taken before (x axis negative) and after (x axis positive) the deals, with 90% confidence intervals in grey. Results suggest no discontinuity, either in level or in trend, at the deal year.

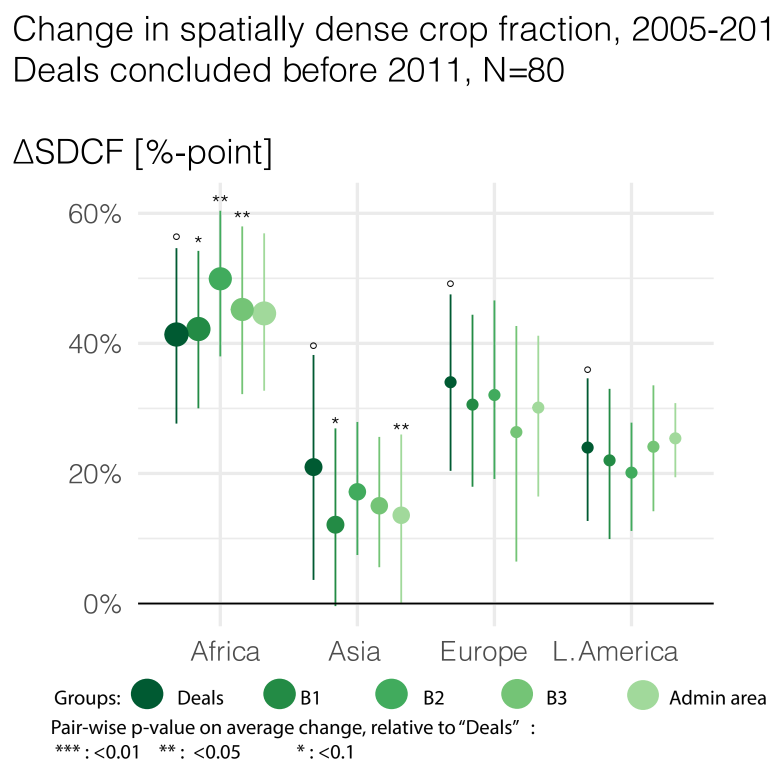

Fig. S5. ***Split Sample Analysis***: Change in the proportion of cropland that is intensively ($> 70\%$) cultivated for the subset of deals concluded before 2011. Results are nearly identical to the outcome presented in Figure 2C of the main document, where the full sample of deals were considered. These results do not support the hypothesis that the absence of effect of LSLA's on the spatial consolidation of cropland is due to a delayed effect (see Materials and Methods in main document).

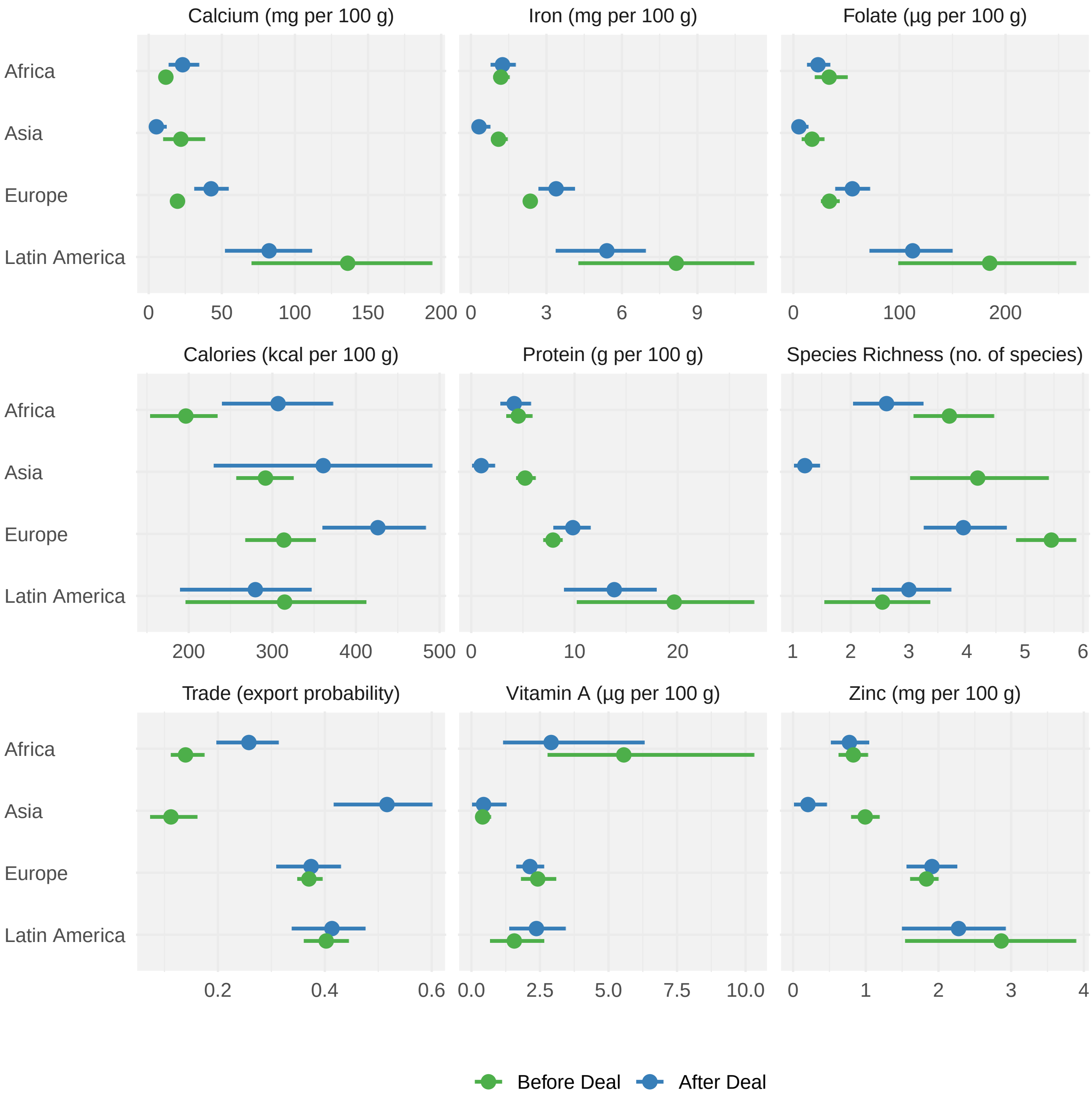

Fig. S6. ***Average effect of intended crop transitions by deal on crop diversity, nutrient production and trade likelihood***. Outcome C_{vt} associated with nutrient v in period t (before or after the deal) was obtained as $C_{vt} = \frac{1}{N_d} \sum_{c,i} F_{cit} X_{civ}$, where F_{cit} is the fraction of harvested area attributed to crop type c in deal i and X_{civ} the associated nutrient content (see Materials and Methods in main document).

Fig. S7. ***Crop category transitions by area of acquired land***: The average crop compositions of deals by continent were computed by weighing the fraction of harvested area F_{cit} attributed to each crop category c for each deal i, by the total surface area A_i of the deal: $F_{ct} = \frac{\sum_i A_i}{\sum_i A_i F_{cit}}$

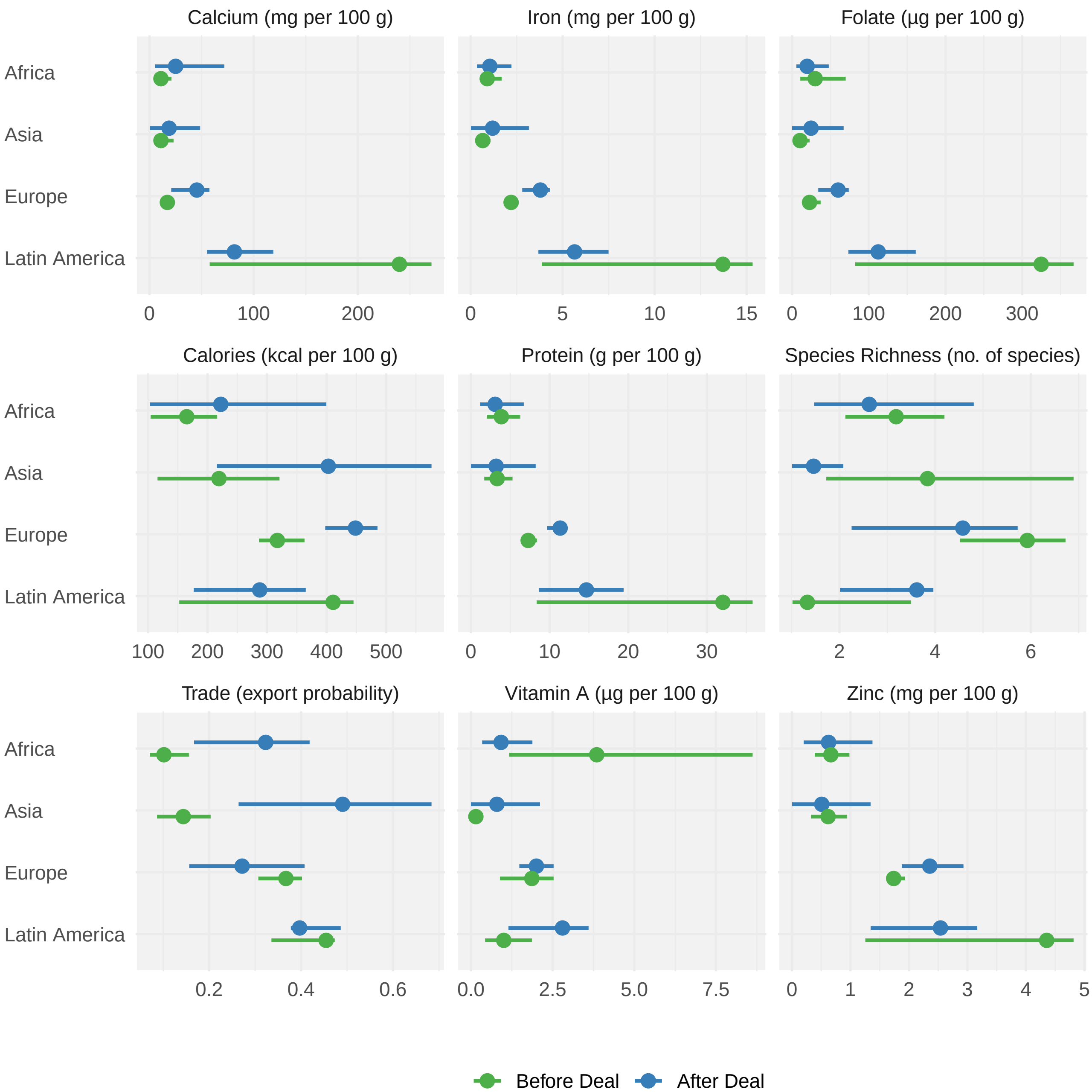

Fig. S8. ***Average effect of intended crop transitions by area of acquired land, on crop diversity, nutrient production and trade likelihood***: Outcome C_{vt} was obtained by weighing the contribution of each deal by its land surface area A_i: $C_{vt} = \frac{1}{\sum A_i} \sum_{c,i} A_i F_{cit} X_{civ}$, where F_{cit} is the fraction of harvested area attributed to crop type c in deal i and X_{civ} the associated nutrient content (see Materials and Methods in main document)

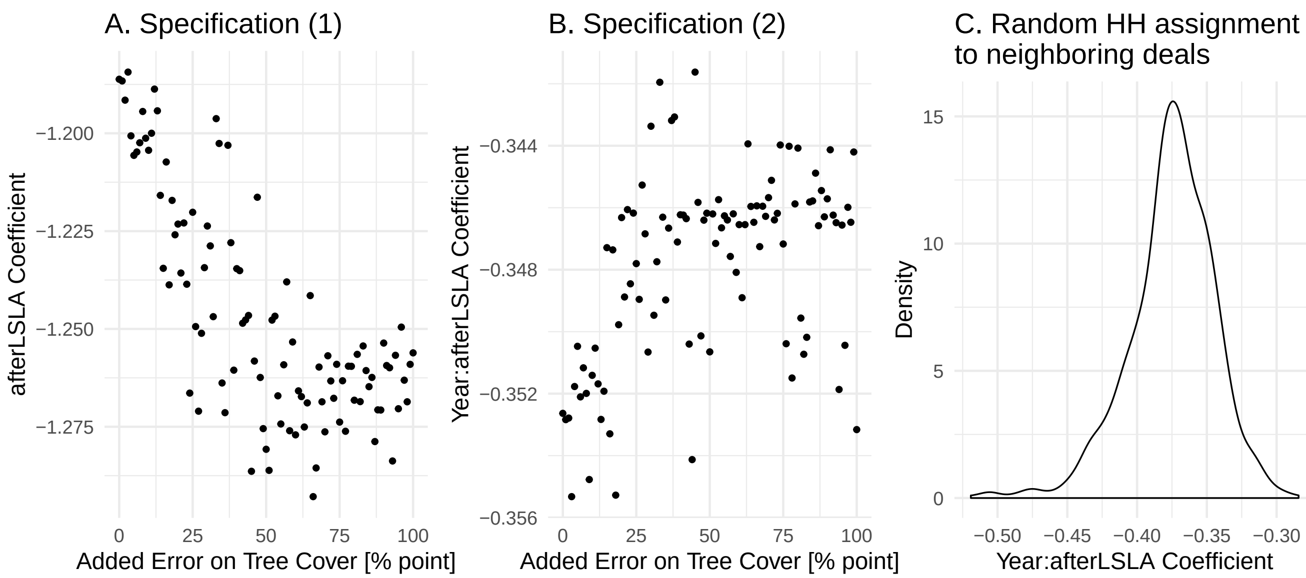

Fig. S9. ***Robustness checks on Household Analysis*****.*****Panels A and B:*** Random errors drawn from independent uniform distributions were added to tree cover observations to simulate detection errors. The maximum amplitude of the added error (i.e. the bounds x of the considered uniform distributions $U[-x, x]$) are displayed on the x-axis. The corresponding regression estimates (namely the *afterLSLA* and *afterLSLA:Year* coefficients for regressions (1) and (2) on Table S6) are displayed on the y-axis. They lie within 1% of the original regression coefficient and display identical statistical significance (namely $p < 0.01$ and $p < 0.1$, respectively). This suggest that the results presented in Table S6 are robust to errors in Tree Cover observations. ***Panel C:*** Household within deals with overlapping control areas (B1-B3) are randomly assigned to the corresponding deals in order to account for the random (and unknown) spatial displacement associated with DHS clusters. The panel displays the histogram of the estimated *afterLSLA:Year* coefficients for regression (2) in Table S6 (N=1000 repetitions). All results are statistically different from zero ($p < 0.1$) and within a 50% error margin of the original regression coefficient.