Delinking land rights from land use: Certification and migration in Mexico Online Appendix

Alain de Janvry, Kyle Emerick, Marco Gonzalez-Navarro and Elisabeth Sadoulet

In this appendix we provide additional details on construction of some of the data used in the analysis. We also provide additional analysis supporting some of the main findings in the paper.

Progresa Data

Household level migration was taken from the 1998-2000 fall versions of the ENCEL survey. The survey was conducted each fall from 1998-2000 in the 506 localities that were part of the experimental evaluation of Progresa. Since no ejido identifiers were included in these data, we matched the 506 localities to ejidos using a spatial join in ARCGIS. We only observe the coordinates of the centroid of each locality and therefore match localities to ejidos if the center of the locality is located inside the boundaries of the ejido. The digital maps of all ejidos certified from 1993-2006 were obtained from RAN. The spatial merge resulted in 234 of the localities falling into one of 219 different ejidos.[1] The number of households from the 1998 survey that fell inside ejidos as a result of this process is 13,212. Another 4,893 households were removed from the sample as a result of being in ejidos that were certified before 1997. Since permanent migration is being measured, trends in migration are unlikely to be the same in ejidos certified prior to 1997 as those certified later. These ejidos are removed for this reason. It is also important to note that the spatial matching approach does not result in a perfect match between households and ejidos. It is possible that while the centroid of a locality falls into a particular ejido, the outskirts of the locality fall into a different ejido. This is more likely to be an issue in

[1] This number is roughly consistent with half of Mexico's land being in ejidos. The large number of localities that were not matched to ejidos is therefore not a concern. The matching rate of 46% is actually in line with 50% of land being in ejidos.

localities that are large. We used census population data to construct the ratio of the population of the locality to the number of ejidatarios in the matched ejido. The matching is more likely to be inaccurate when the locality is large relative to the ejido. We therefore retained only the 200 localities with the lowest values of this metric. This amounted to removing an additional 742 households from the sample. The total number of ejidos in the sample is 127.

1991 and 2007 Ejido Census

The 1991 and 2007 ejido censuses consist of a set of 28,752 ejidos that were surveyed in both 1991 and 2007. We were unable to obtain the name of each ejido due to confidentiality concerns. Further, the 2007 census did not contain information on the time of completion of Procede. A matching process was therefore necessary to make these data usable. The key information used were the state, municipality, and name of the locality where the majority of the ejidatarios live. We used this information along with some common key variables between the census data and the GIS database from RAN to match ejidos based on a 4-step process:

1. There were 22,473 ejidos for which the locality where a majority of the ejidatarios live is located inside the boundaries of the ejido. For these ejidos we were able to use our spatial merge between localities and ejidos to identify the corresponding ejido in the GIS database. There are of course numerous instances where the boundaries of an ejido contain more than one locality centroid. We were unable to include these ejidos in this matching round. This round matched a total of 14,128 ejidos.

2. The second round of matching is meant to partially correct for the fact that matching localities to ejidos in the previous step using only the centroid of the locality is imperfect. The reason for this is that the centroid of the locality could fall outside of the boundaries of the ejido even if there is substantial overlap between the locality and ejido. Further, ejidos with multiple disjoint patches of land pose problems to matching based on locality centroids and ejido boundaries. The distance between the locality centroid for each unmatched census ejido and the center of each unmatched ejido from the GIS database was calculated using a simple distance calculation in ARCGIS. An ejido from the GIS data was matched to an ejido from the census data if the locality where the majority of the ejidatarios live was the closest locality to the center of the ejido. Since this match is not perfect, we attempt to minimize errors by only retaining matches where the percentage difference between the number of ejidatarios in the

1991 census and the GIS database was between -46.8% and 29%.[2] This round generated an additional 1,787 matches.

3. In this round we considered the remaining unmatched ejidos for which the locality where the majority of the ejidatarios live is located inside the boundaries of the ejido. We defined a potential candidate match from the GIS database as an unmatched ejido that was located in the same state and municipality. For each of these potential matches we considered 4 metrics of comparison. The first was the similarity between the name of the locality where the ejidatarios live and the name of the ejido in the GIS database.[3] We generated a spelling similarity index using a combination of the COMPARE and SPEDIS functions in SAS. A match was identified for sufficiently low values of this index. The second metric was the distance between the centroid of the locality and ejido. The ejidos were considered to match if the distance was less than 5.1 kilometers.[4] The third metric was the number of ejidatarios. A match was determined using the same cutoffs as in the previous round. The final metric was the difference between the size of the ejido (in hectares) in the two datasets. The percentage cutoffs were -32.4 and 41.6. We required at least two of these criteria to be satisfied to identify a match between the ejidos. For each census ejido we selected the ejido from the GIS database which matched on the most of these criteria (from 2 to 4). In order to break ties we used the percentage difference in the number of ejidatarios. This round generated a total of 1,878 matches.

4. The fourth round of matches considers the census ejidos where it was stated that the locality where the majority of ejidatarios live is *not* inside the boundaries of the ejido. We used a similar process as in the previous round with only two modifications. First, similarities between the name of the locality and the ejido were not used. Second, the distance requirement was relaxed to 8.6 kilometers (25th percentile). This round generated 1,920 matches.

Data on PROCAMPO beneficiaries

We use publicly available data from Mexico's secretariat of agriculture (SAGARPA) on the farm support program PROCAMPO to analyze land use changes after

[2]These numbers were chosen as the 10th and 90th percentiles of the percentage difference from the ejidos matched in the previous round.

[3]It is common for ejido names to be the same as locality names in Mexico.

[4]This value was chosen since it was the 10th percentile in the list of candidate matches.

Procede. PROCAMPO entitles farmers that cultivated land during the period from 1991-1993 decoupled payments per hectare cultivated. The first payments were made to approximately 3.2 million beneficiaries during the main spring/summer growing season in 1995.

We observe each claim made by each of these beneficiaries during the period from 1995 to 2012.[5] Overall, the data contain information on 45.6 million support payments over the 18 year period. Each observation consists of the name of the farmer, numeric farmer identifier, the location as defined by state, municipality, and ejido, the crop, and area cultivated. These payments track closely the cultivation patterns of the land cultivated during the base period because the only requirement for receiving payment was cultivation of the land.

Ejidos are not identified numerically and the data are not spatially referenced. This necessitated a name match where we were able to successfully match farmers from 19,409 different ejidos to our data on program completion. The ejido-level analysis in Figure 2 of the main text uses 18,437 of the 19,409 ejidos that had observations from both 1995 and 2012.

Additional analysis on land use

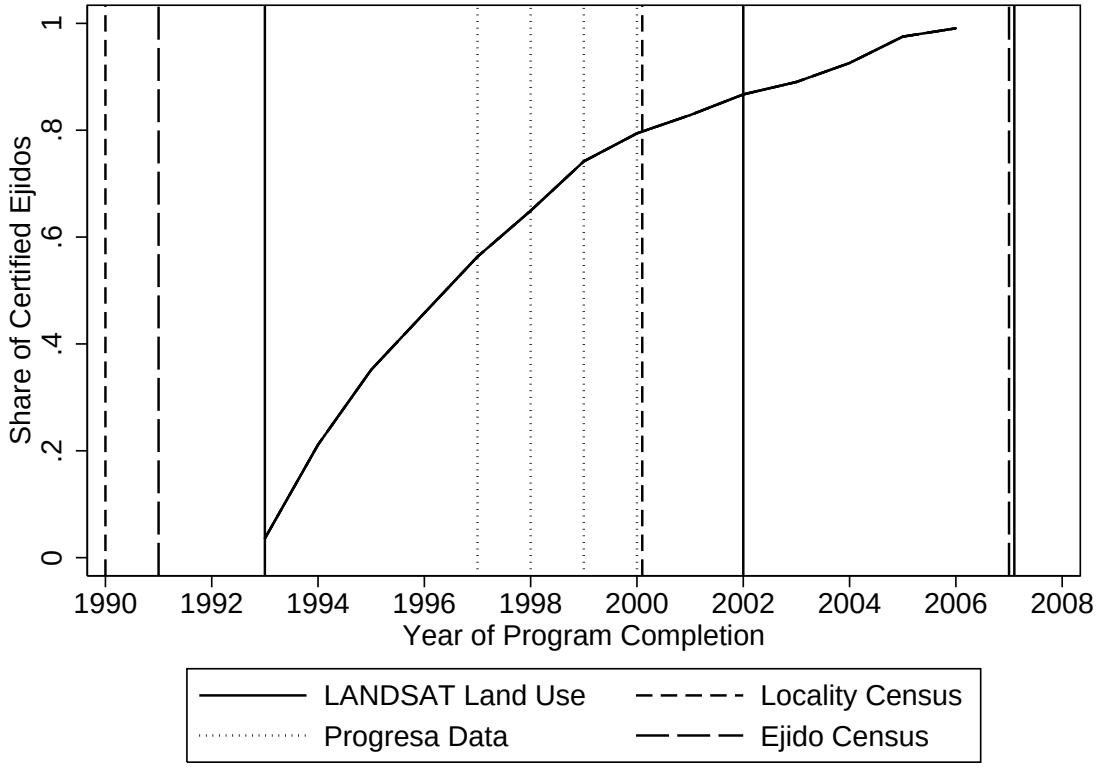
We show that there is a correlation between population changes and cultivated area changes. In order to do this, we consider the overall change in log agricultural land between 1993 and 2007 using the Landsat data. The median change in log of agricultural land in these data is .0001 while the mean is 0.111. To limit the influence of outliers, we use the rank of the ejidos in the distribution of change in cultivated land.[6] The first two columns of Table A2 repeat the fixed effects regression of locality population on whether the ejido has been certified separately for the localities with agricultural land use change below and above the median value. The table shows that the negative effect of certification on population size is much stronger in localities that also saw the largest decreases in agricultural land.

Column (3) shows that localities with the most pronounced declines in agricultural land ($rank$ =0) experienced a decline in population of 9.2% in response to certification, while ejidos with the largest increases in agricultural land saw no significant effect of certification on population.

[5] We don't observe payments for the main growing season during 1998.

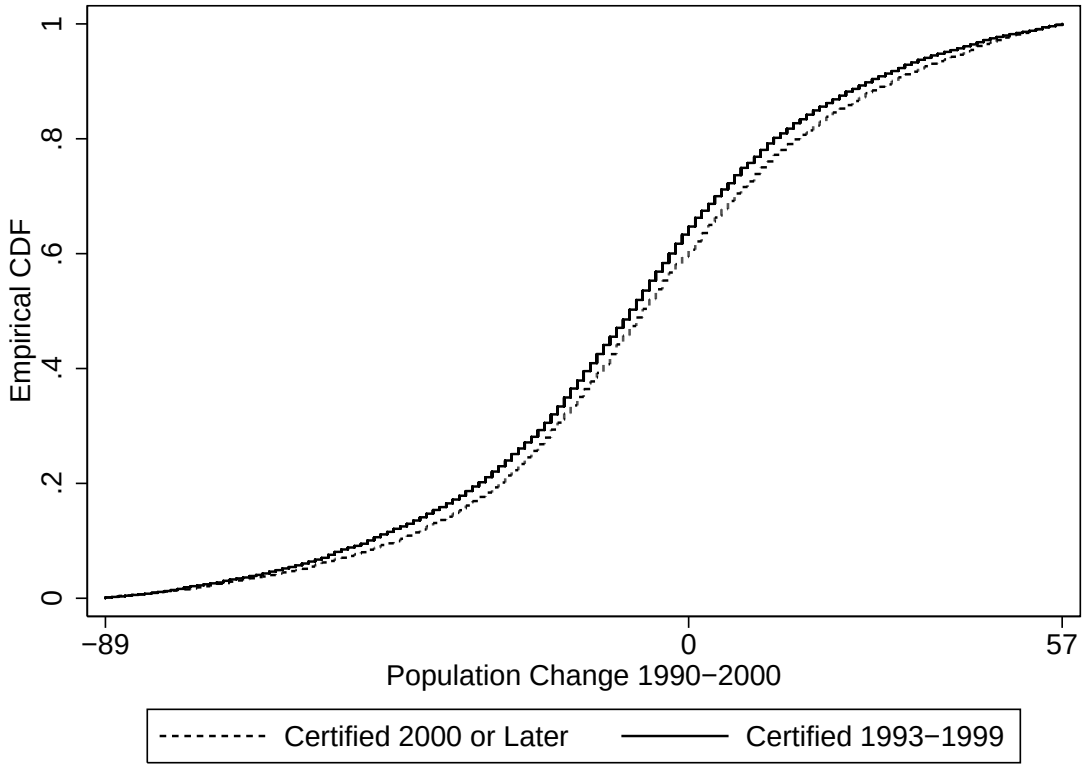
[6] The value of the variable Rank corresponds to the empirical distribution function of the change in the logarithm of agricultural land.

Figure A1: Correspondence between data and rollout of Procede



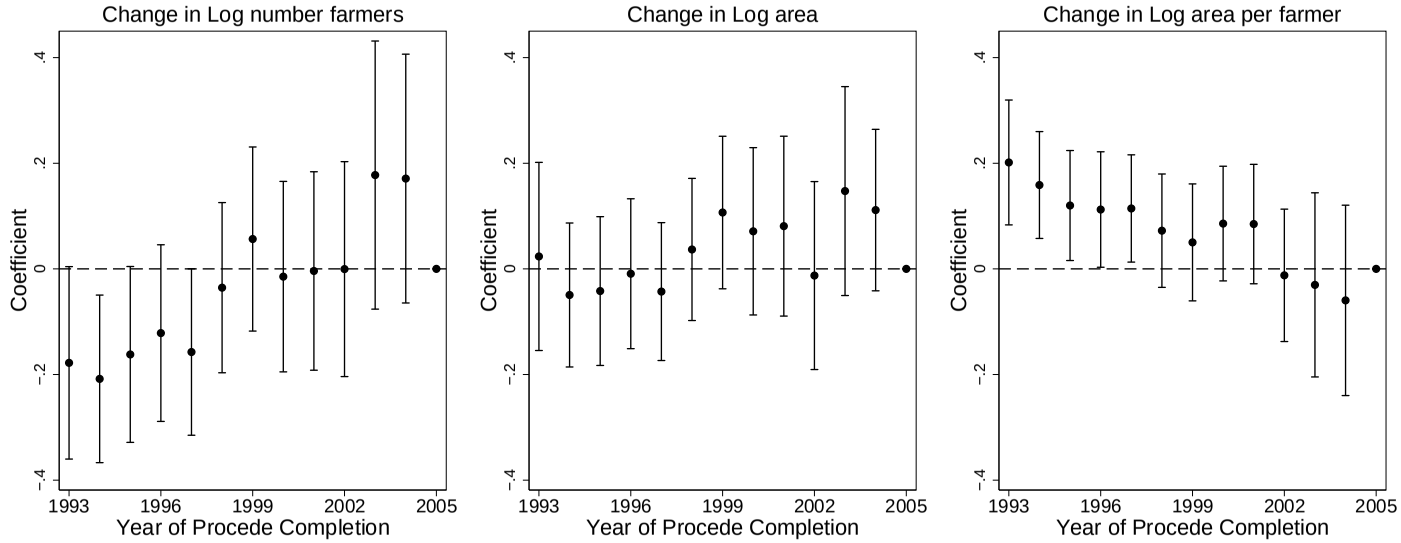
Notes: Figure shows cumulative share of ejidos certified over time. Vertical lines represent observations for each of the datasets used in the analysis. The Progresa ENCEL data are from 1998-2000. Migration recall data were used for 1997. Locality level census data are from 1990 and 2000. Ejido level census data are from 1991 and 2007. LANDSAT land use data are from 1993, 2002, and 2007.

Figure A2: Cumulative distribution of population change, 1990-2000, by certification date



Notes: Figure displays empirical CDF of population change (in levels) from 1990-2000. Data used are the 1990 and 2000 locality-level population censuses.

Figure A3: Analysis of migration and land consolidation for subset of ejidos without any common land at baseline



Notes: Figures show coefficient estimates from changes in log number of cultivators and total area cultivated from1995-2012 on year of Procede completion indicators and state fixed effects. Data are limited to areas where there was no common area prior to Procede. The standard errors are clustered at the municipality level. The omitted category is 2005 and onwards, thus coefficient estimates are relative to ejidos that were completed near the end of the program.

Table A1: Heterogeneous effects of certification on migration

	Progresa Households Matched to Ejidos					
	(1) Has Migrant	(2) Has Migrant	(3) Has Migrant	(4) Has Migrant	(5) Has Migrant	(6) Has Migrant
Certified	0.0134	0.0074	0.0298***	0.0312***	0.0234***	0.0266***
	(0.0087)	(0.0086)	(0.0097)	(0.0101)	(0.0076)	(0.0079)
Certified*Ejido Had Boundary Problems in 1991	0.0164	0.0285*				
	(0.0142)	(0.0145)				
Certified*High Maize Yield Municipality			-0.0232*	-0.0256**		
			(0.0124)	(0.0126)		
Certified * Land per Adult ¿ Median in Ejido (1997)					-0.0187*	-0.0265**
					(0.0097)	(0.0116)
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes
Ejido Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Effects*Ejido Had Boundary Problems in 1991	No	Yes	No	No	No	No
Time Effects*High Maize Yield Municipality	No	No	No	Yes	No	No
Time Effects*Land per Adult > Median in Ejido	No	No	No	No	No	Yes
Mean of Dep Variable	0.057	0.057	0.056	0.056	0.056	0.056
Number of Observations	20929	20929	24372	24372	24372	24372
R squared	0.059	0.060	0.058	0.058	0.058	0.058

Standard errors that allow for clustering at the ejido level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. Data include observations on all households in ejidos that completed the Procede process after 1996. All regressions are linear probability models. Dependent variable = 1 if the household had a migrant leave during the year or any previous sample year. Certified indicator = 1 if ejido was certified at the start of the year. All regressions include indicator for landholdings per adult being above median in ejido, age of household head, indicator for female household head, and number of males between 17 and 30 as controls. *Ejido had boundary problems* is an indicator for ejidos that reported having boundary problems during the 1991 ejido survey.

Table A2: Population regressions by change in agricultural area

	Rank>0.5	Rank<0.5	All
	(1)	(2)	(3)
	ln(Population)	ln(Population)	ln(Population)
Year=2000	-0.2285***	-0.1936***	-0.1765***
	(0.0143)	(0.0195)	(0.0239)
Certified 1993-1999*Year=2000	-0.0230	-0.0760***	-0.0924***
	(0.0183)	(0.0232)	(0.0292)
Rank of Ag Change * Year=2000			-0.0705*
			(0.0368)
Rank of Ag Change * Certified 1993-1999 * Year=2000			0.0876*
			(0.0461)
Ejido Fixed Effects	Yes	Yes	Yes
Mean of Dep Variable	4.240	4.324	4.278
Number of Observations	15200	12420	27624
R squared	0.035	0.041	0.038

Dependent variable in all regressions is log of locality population. Standard errors that allow for clustering at the ejido level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***,5% **, and 10% * levels. Data come from the 1990 and 2000 locality population censuses. Localities located in ejidos with no agricultural land during either 1993 or 2007 are excluded from the regressions, thus explaining the difference in observations from Table 2. The first column limits to localities in ejidos that experienced above the median change in log agricultural area from 1993-2007. The second column limits to localities in ejidos that experienced below the median changes. The final column is for all localities in ejidos that had nonzero agricultural land area in both 1993 and 2007.

Table A3: Direct effect of Progresa on household-level migration

	Migration in:		
	(1)	(2)	(3)
	1998	1999	2000
Progresa treatment locality	0.0224***	0.0221*	0.0124
	(0.0082)	(0.0128)	(0.0136)
HH is Landholder	0.0196***	0.0138	-0.0064
	(0.0073)	(0.0102)	(0.0100)
Number Males 17-30 in HH	0.0096	0.0148*	0.0175*
	(0.0063)	(0.0081)	(0.0092)
HH head is Female	0.0140	0.0071	0.0205
	(0.0169)	(0.0207)	(0.0215)
Age of HH Head	0.0010***	0.0017***	0.0017***
	(0.0002)	(0.0003)	(0.0003)
Mean of Dep Variable	0.041	0.061	0.078
Number of Observations	3509	3107	3205
R squared	0.013	0.014	0.011

Data are for all poor households that were eligible to receive Progresa payments. Dependent variable = 1 if the household had a migrant leave during the year or any previous sample year. Standard errors that allow for clustering at the locality level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A4: Relationship between Procede and pre-program migration

	Progresa Households Matched to Ejidos, Pre-Program Period			
	(1)	(2)	(3)	(4)
	Δ Migration,97-98	Δ Migration,97-99	Δ Migration,97-00	Migration,97-00
Procede Completed in 1999	-0.0011			
	(0.0113)			
Procede Completed in 2000	-0.0040	-0.0087		
	(0.0110)	(0.0092)		
Procede Completed After 2000	-0.0131	-0.0102	0.0015	
	(0.0090)	(0.0086)	(0.0046)	
Year Procede Completed (0/1)				0.0018
				(0.0150)
Year Before Procede (0/1)				-0.0021
				(0.0107)
2 Years Before Procede (0/1)				-0.0015
				(0.0089)
Time Fixed Effects	No	Yes	Yes	Yes
Ejido Fixed Effects	No	No	No	Yes
Mean of Dep Variable	0.022	0.020	0.018	0.050
Number of Observations	111	187	225	406
Number of Ejidos	111	94	76	127
R squared	0.047	0.019	0.002	0.774
Pvalue of joint test	0.190	0.493		

Standard errors are reported in parentheses. Robust standard errors are used in column 1. In columns 2-4, standard errors are clustered at the ejido level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. The dependent variable in columns 1-3 is the change in ejido migration rate. The dependent variable in column 4 is the ejido migration rate. Both regressions are for the pre-treatment period. Columns 1 is for for 1998. Column 2 is for 1998-1999. Column 3 is for 1998-2000. Column 4 is for 1997-2000.

Table A5: Regressions of attrition on certification status and household covariates

	(1) Attrition
Certified	-0.003
	(0.025)
HH is Landholder	-0.043***
	(0.010)
Number Males 17-30 in HH	0.005
	(0.004)
HH Head is Female	0.030**
	(0.012)
Age of HH Head	-0.000
	(0.000)
Ejido Fixed Effects	Yes
Time Fixed Effects	Yes
Mean of Dep Variable	0.112
Number of Observations	12895
R squared	0.115

Standard errors that allow for clustering at the ejido level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. Data are for all households that were surveyed in the Fall 1998 ENCEL survey. Observations are from 1999 and 2000. Dependent variable = 1 if household did not have survey completed. Certified indicator = 1 if household had a certificate at the start of the year. 446 households attrited in 1999 but not in 2000. 331 households attrited in both 1999 and 2000. 554 households attrited in 2000 but not in 1999.