Measuring the Impact of Land Titling on Investment via Satellite

CEGA Research Retreat

Luke Sanford, Jen Burney, Ran Goldblatt
UC San Diego, School of Global Policy and Strategy

Klaus Deininger
The World Bank

BIG PIXEL
INITIATIVE AT UC SAN DIEGO

What we are doing here?

A CEGA-Inspired Project!

- Panel at Geospatial Analysis for International Development:
  - The truth isn't necessarily on the ground
  - We're still figuring out what's possible (and not yet possible)
  - How to use different degrees of freedom?
- Can we use RCTs on the ground to test the possibilities/limits of Remote Sensing for impact evaluation?

Two Reasons to Stay for the Rest

Substantive:
Does formalization of landholdings lead to greater on-farm investment?

- Yes: Modest Impacts
- Meh: SSA, Heterogeneity

We have some (preliminary) evidence towards yes in a few different ways.

Methodological:
Can remote sensing help us conduct longer-run, ex-post evaluations?

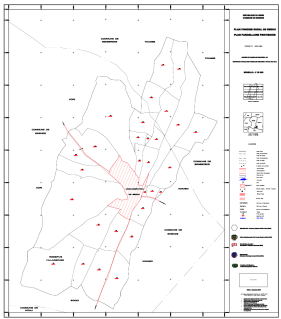
- Night Lights, Built-up Area, Yields
- Us: Drivers of LUC @ farm scale

Cool tools to learn more from the information that already exists.

What are we Actually Doing?

World Bank project (Plans Fonciers Ruraux) to formalize landholdings across Benin

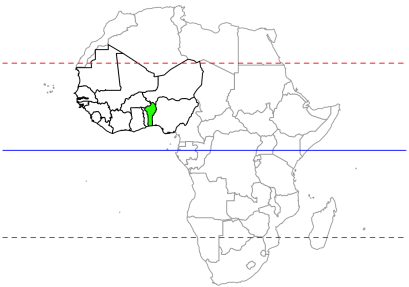
- National scale program
- 2009-2011 (mostly)
- Randomized at village level
- Village-level: conflict resolution, plot demarcation, title issuance
- $N_{village}$ = 300 and n_{plot} = 70,000
- HH survey from 2011 (WB)



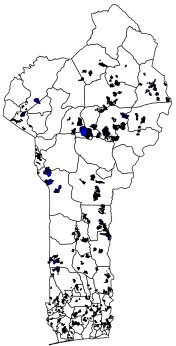
Look for fingerprints of impact in remotely-sensed vegetation indices. Do we see what the conventional impact evaluation saw?

Quick PFR Summary

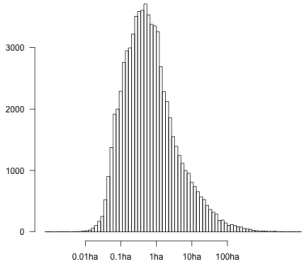
Benin



PFR Villages



Plot Size Distribution



Why is this a Good Project to Use? (1)

	Obs.	Control		ITT	
		Mean	Std. dev.	Coeff.	s.e.
Parcel has clear borders[†]	6,094	0.061		0.270***	(0.02)
Panel A: Tenure security					
Fear of land loss during fallow[†]	6,094	0.116		0.007	(0.01)
Land conflict[a†]	6,094	0.052		-0.009	(0.01)
Perceived land rights					
- *A HH member will inherit land*	3,582	0.829		-0.007	(0.02)
- *The HH head can lend/rent-out/give the parcel*	3,582	0.731		0.017	(0.02)
- *The HH head can pledge land*	3,582	0.719		0.000	(0.02)
- *The HH head can sell land*	3,582	0.554		0.040*	(0.02)
Panel B: Investment and land transfers					
Investment in tree planting[a†]	6,094	0.040		0.017**	(0.01)
Started fallowing parcel[a†]	6,094	0.010		0.004	(0.00)
Parcel is rented in/out[†]	6,094	0.147		-0.014	(0.01)
- *rented in*	6,094	0.082		0.002	(0.01)
- *rented out*	6,094	0.065		-0.016*	(0.01)
Panel C: Agricultural activities[b]					
Total land size (ha)	2,972	6.236	13.827	0.320	(0.54)
Participation in agriculture[†]	2,972	0.902		0.006	(0.01)
Share of land size cultivated	2,675	0.538	0.323	-0.001	(0.01)
Panel D: Agricultural production[c]					
Type of crop cultivated					
Cereals[†]	6,094	0.505		0.001	(0.02)
Pulses[†]	6,094	0.150		0.003	(0.01)
Roots and tubers[†]	6,094	0.245		-0.005	(0.02)
Vegetables[†]	6,094	0.052		-0.004	(0.01)
Cash crops[d†]					
- *annual*	6,094	0.037		0.001	(0.01)
- *perennial*	6,094	0.067		0.026**	(0.01)
Inputs					
- *farm labor supply (person-days/ha)*	3,994	202.854	261.071	1.690	(9.88)
- *fertilizer/high-yield seeds*[†]	3,994	0.272		0.018	(0.02)
Output					
- *total output (Log USD)*	3,677	6.135	1.358	-0.043	(0.06)
- *yield (Log USD of total output/ha)*	3,677	6.379	1.064	0.023	(0.05)

A few percent increase in tree planting and cultivation of perennial (e.g. trees) and annual cash crops.

Why is this a Good Project to Use? (2)

While demarcation activities do lead to an increase in long-run investments, they do not generate an initial increase in agricultural output, farm yields (measured as the log of the value harvested per hectare), or the use of productivity-enhancing inputs such as labor, fertilizer, or improved seeds. Given that it takes more than one year for these long-term investments to bear fruit, it is unsurprising that demarcation does not generate productivity gains at this early stage in implementation.[32]

[32] The gestation period for cashew and oil palm trees, for example, is at least four to five years.

The More Nuanced Truth

- Obvious Pros: An RCT with a few small significant effects over a short time period, and a lot of other null findings.
- Less Obvious Pros: No geographic control information, and no baseline! So RCT is irrelevant.
  - Geospatial synthetic control opportunity
  - Pretty independent cross-validation

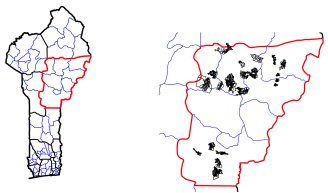
Potential to open up a lot of other un-evaluated interventions for observational analysis.

Goals

1. What effects do we measure remotely?
    - Production time scale is slow compared to stagger and evaluation
    - Learning curves may exist, weather definitely exists
    - Even immediate changes might take a few seasons to be "visible"
2. Test-drive geospatial synthetic controls:
    - Generate a synthetic control for every treatment observation
    - Create random draw of plots/pixels from all non-treated areas
    - Derive weights for comparisons to generate a control that matches the treatment observation's pre-intervention levels and trends

How the Sausage is Made (Methods)

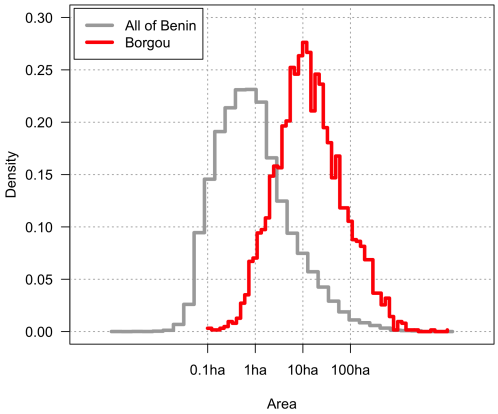
Test Region: Borgou Department



Borgou Plot Stats:

- 3132 treated plots
- PFR rolled out from late 2009 to early 2011
- 1.5 % female titleholders

Borgou Plots are Larger:



Outcomes?

Many types of investment (pathways) we'd like to measure:

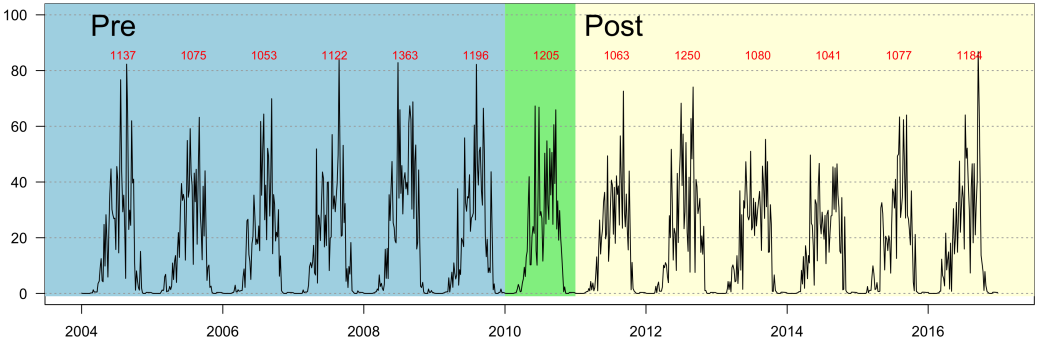
- Fertilizer, HYV use
- Irrigation / water management
- Diversification, high-value crops
- Fallowing, rotation practices

These map non-uniquely to what we can measure @ scale of 1-30m:

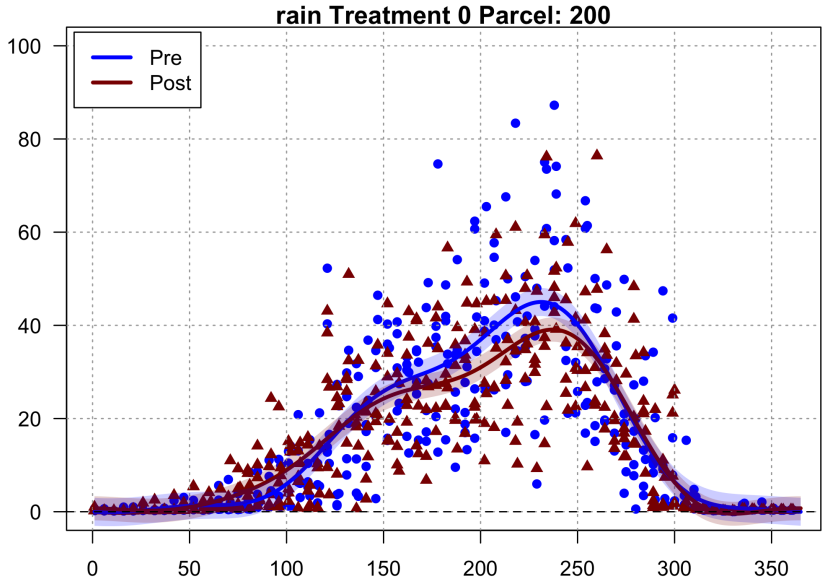
- Productivity (for same crop)
- Seasonal production cycle
- Relationship with precipitation
- Fallowing, rotation practices

Here we focus on perennial and counter-seasonal production.

Seasonality



Decrease in Rainfall



Outcomes

All on a Seasonal Basis:

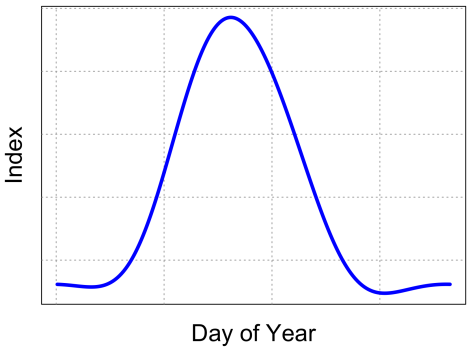
- Predicted (fit) Max VI value
- Day of Max Predicted VI value [1:365]
- Is Max Predicted VI in dry season? (binary)
- Average VI, end of main season
- Average VI, dry season (1)
- Average VI, dry season (2)

Data Sources:

- Landsat 7 8-day TOA Corrected Composites in GEE
  - Extract time series for all treated plots (average) and control points (pixels)
  - Generate NDVI, EVI, SAVI, LSWI, GCVI
  - Get dry season metrics
  - Fit seasonal pattern in pre- and post- treatment period
- CHIRPS (5-day) precip (seasonality)

Fitting Annual Cycles

Stack annual values:



Fit arbitrary sinusoid:

$$y = I_0 + A_1 sin\left(\frac{2\pi J}{365}\right) + A_2 cos\left(\frac{2\pi J}{365}\right) + B_1 sin\left(\frac{4\pi J}{365}\right) + B_2 cos\left(\frac{4\pi J}{365}\right) + C_1 sin\left(\frac{6\pi J}{365}\right) + C_2 cos\left(\frac{6\pi J}{365}\right) + D_1 sin\left(\frac{8\pi J}{365}\right) + D_2 cos\left(\frac{8\pi J}{365}\right)$$

Control Strategy

- No control plots
- No control villages
- Basically observational data
- What would a good control look like?

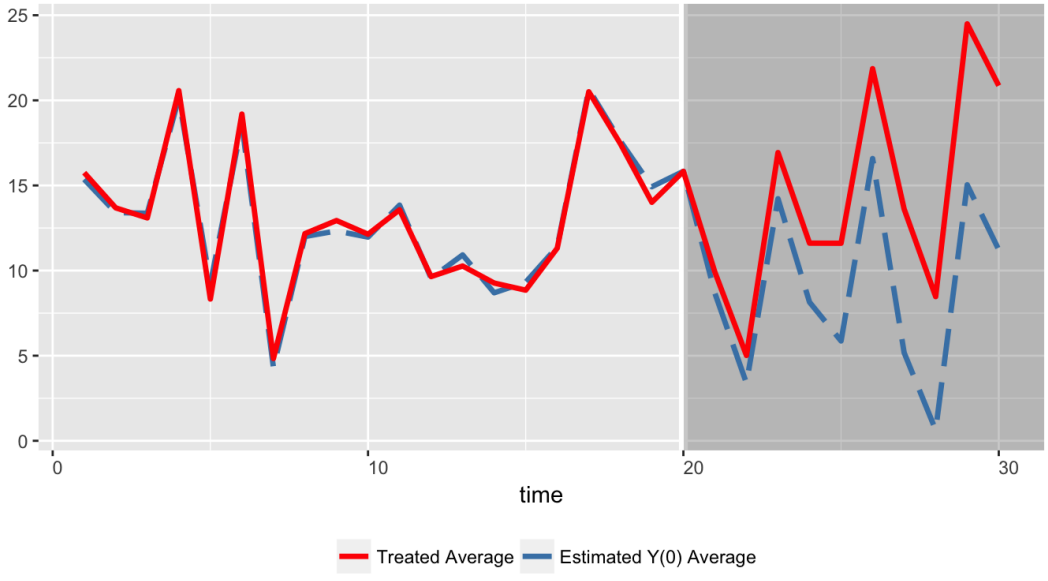
Control Strategy

- Would match a treated unit exactly before treatment
- Synthetic control method **Abadie Diamond and Hainmueller 2010**
- Interactive Fixed Effect method **Bai 2009**
- Generalized synthetic control method **Xu 2017**

Estimation

- Estimate latent factors and factor loadings from control group via IFE
- Find the factor loadings that best predict the treated units in the pre-treatment period
- Impute counterfactuals in post-treatment period based on latent factors and these factor loadings
- An estimator $A\hat{T}T_t$ is $(1/N_{tr}) \sum_{i \in \tau} [Y_{it}(1) - \hat{Y}_{it}(0)]$ for $t > T_0$

Estimation-Example

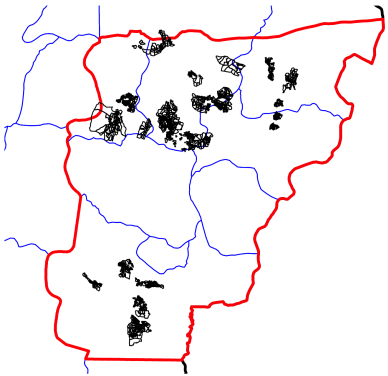


Estimation

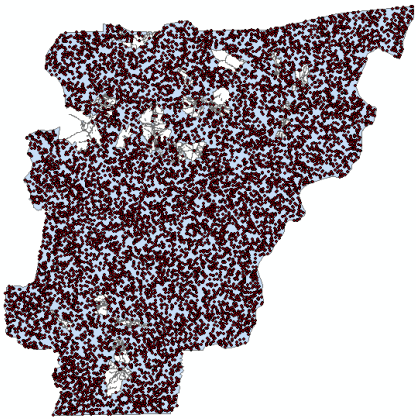
- Treated units are plots (average of pixel values in plot)
- Control units are points from Borgou outside of the treated plots
- Synthetic control is (basically) a weighted average of the untreated points that match a treated plot best

Data

Treated Plots



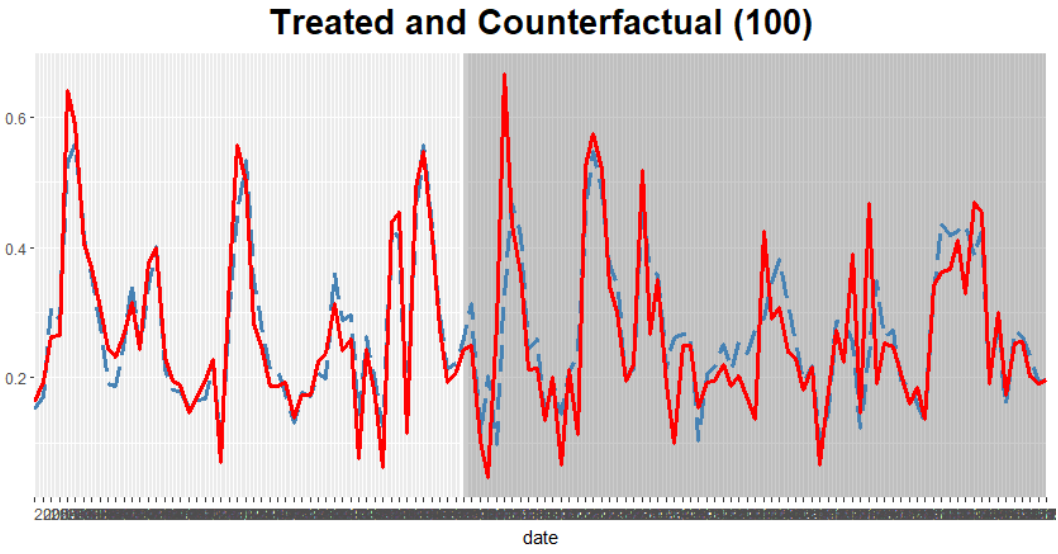
Untreated Points



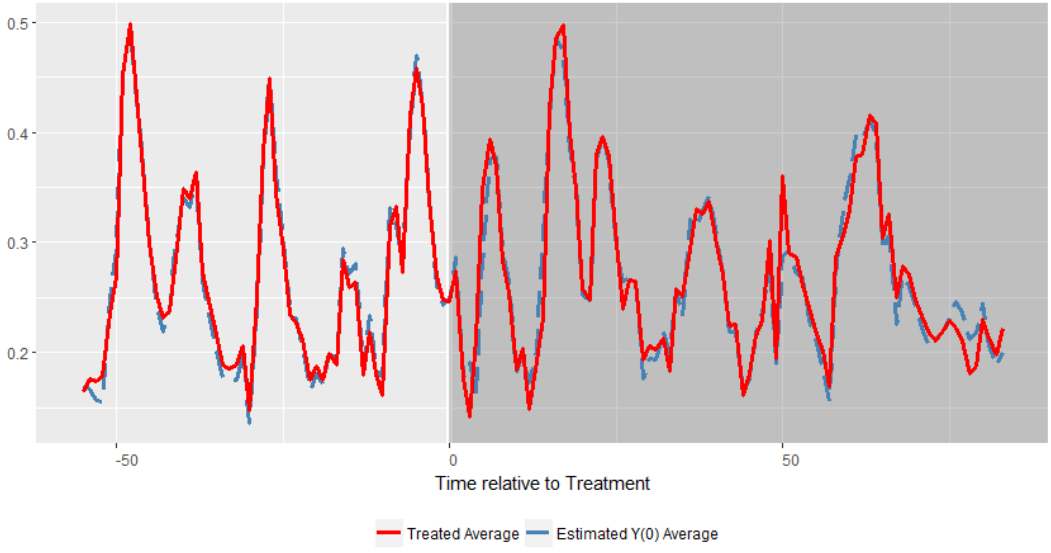
Data

- Begin with 15132 observations, about 250 time periods
- Drop time periods with >50% missingness
- Drop observations with >10% missingness
- Result: 1545 treated, 4881 control, 126 dates from 2005-2017
- Output 3090 observations. Half treated, half imputed control

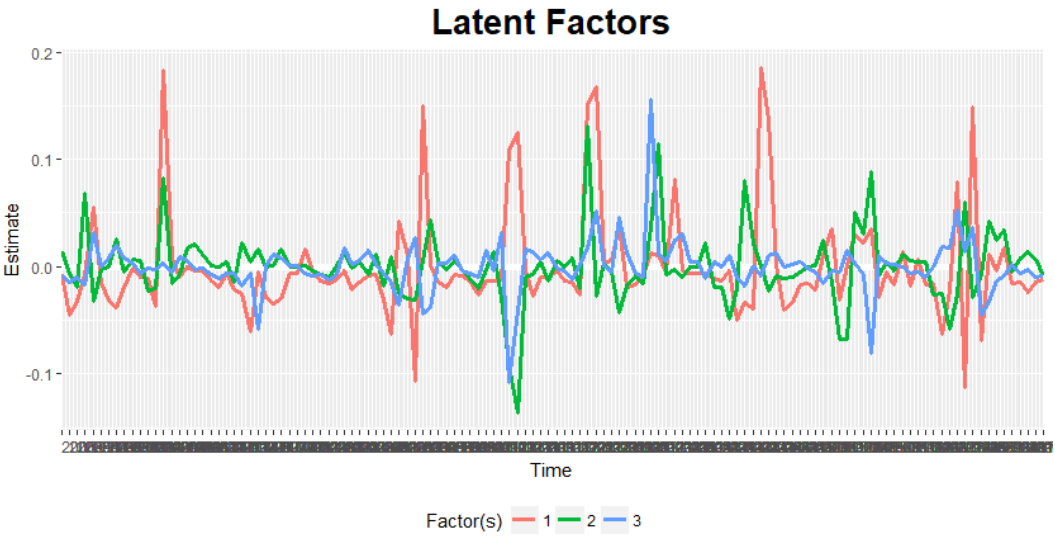
Estimation-One of our examples



Estimation-Whole sample



Factors



Estimation

Difference-in-Differences:

$$y_{it} = \beta T_{it} + s_i + \theta_t + \epsilon_{it}$$

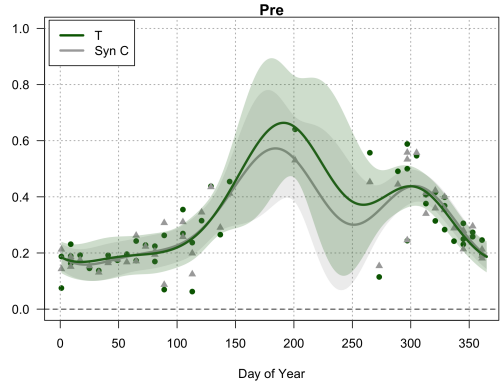
- y_{it}: DVs, I_0 and 1st, 2nd, 3rd Harmonics
- i: plot (either PFR or synthetic control)
- t: pre- or post- PFR
- s_i: plot FE
- θ_t: pre- post- FE

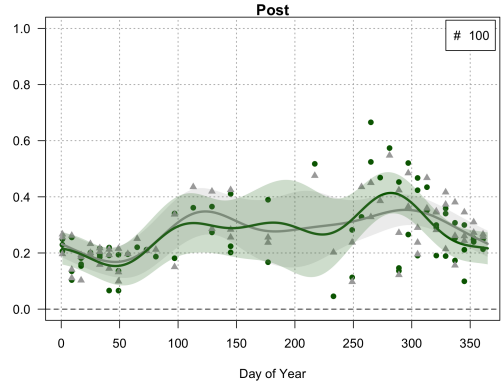
Standard Errors

- They are incorrect at the moment–don't pass-through uncertainty from synthetic control method
- Also don't account for spatial autocorrelation or village clustering

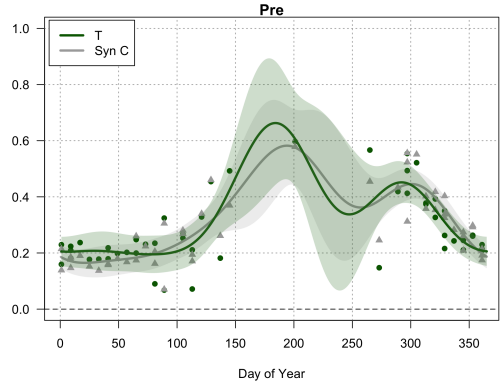
Results

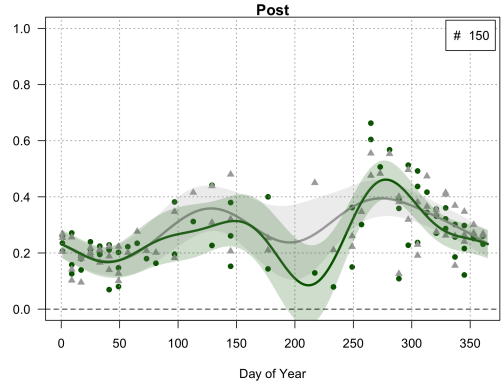
Output of matching and harmonic fitting





And a few more for results





Hypothesis Tests

Table: Preliminary Results

	Dependent variable:					
	NDVI					
	(Max)	(Day of max)	(max in dry season?)	(end of main)	(dry season)	(first 4 months)
post:treatment	−0.095***	41.126***	−0.001	0.001	0.003***	0.012***
	(0.003)	(1.403)	(0.005)	(0.001)	(0.001)	(0.001)
Constant	0.536***	213.664***	0.034***	0.326***	0.223***	0.184***
	(0.001)	(0.702)	(0.003)	(0.001)	(0.0003)	(0.0003)
Observations	6,180	6,180	6,180	6,180	6,180	6,180
R^2	0.144	0.122	0.00001	0.0002	0.002	0.056
Adjusted R^2	0.144	0.122	−0.0002	0.0001	0.002	0.056
Residual Std. Error (df = 6178)	0.101	47.764	0.180	0.036	0.023	0.021
F Statistic (df = 1; 6178)	1,037.128***	859.060***	0.060	1.470	13.942***	368.897***

Note: *p<0.1; **p<0.05; ***p<0.01

Comparison to WB Survey Data

- Lower max NDVI (Perennials?)
- Day of max NDVI is later (Land improvements?)
- No change in whether the date of max NDVI is in the dry season
- No change in NDVI at end of main season
- Higher dry season NDVI (Improvements/perennials)
- Higher NDVI in first four months of the year (Improvements/perennials)

Conclusions

Substantive Results Summary

- Some support for the theory that land titling leads to long-term improvements to land
- This appears as more growing in the dry and off seasons, and/or increase in perennial crops.

Feasibility and Expansion

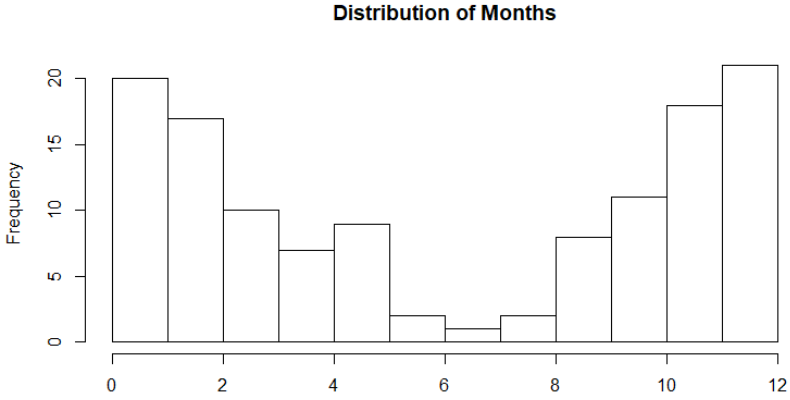
- Feasibility
  - Limits of data extraction from Google Earth Engine
  - Memory and computational intensity of synthetic control method
- Next Steps
  - Different indices in the same region
  - Do different indices work better in different regions?
  - Are the control points farmland?

Bigger Questions

- Further applications of geospatial synthetic control method?

Remaining Methods Questions

- The fit isn't perfect in pre-treatment periods (do we need more data?)
- Missing observations during the rainy season–how does this affect the matching process



Remaining Methods Questions

- What is the right way to measure shift in production?
- If productivity had increased or a shift to perennials had occurred, how would we observe it?

References

Abadie, Diamond and Hainmueller
Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program
Journal of the American Statistical Association
2010

References

Jushan Bai
Panel Data Models with Interactive Fixed Effects
Econometrica
2009

Yiquing Xu
Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models
Political Analysis
2017

Thank you!

lcsanford@ucsd.edu; jburney@ucsd.edu