Measuring the Energy Savings from Tree Shade

Joe Maher
University of Maryland, Department of Agricultural and Resource Economics, College Park, MD 20742, jmaher@arec.umd.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC , August 4-6, 2013.

Abstract: The energy savings from tree shade coincide with peak electricity demand during summer months, creating an opportunity for utilities to use tree protection policies as demand side management tools. We apply a quasi-experimental research design to identify the change in residential energy caused by tree removals using three unique micro-level datasets from Gainesville, Florida. These datasets include (i) a twelve year panel of monthly household electricity billing data for 30,000 homes serviced by Gainesville Regional Utility, (ii) city permit data that identify the timing and location of tree removals, and (iii) property appraisal data detailing structural building characteristics for each home. Results of a difference-in-difference model suggest that removing mature trees in urban setting significantly increases residential energy use. After a tree removal, households experience a 3 percent increase in average monthly utility consumption across the year. The treatment effect is largest during summer months, with an average electricity increase of 4 to 5 percent following a tree removal.

Introduction

Shade trees serve as effective demand-side management tools by cooling homes during times of peak-demand for air conditioning. In warm climates air conditioning accounts for up to 40% of residential electricity during summer months. During the past decade, utilities operating in the U.S. sunbelt have steadily invested in expensive "peaker plants" accommodate peak demand for air conditioning, with costs typically passed through to consumer utility bills. One alternative policy could involve investments in "green infrastructure". Trees shade buildings and cool the interior of homes, moderating the electricity demand shocks caused by extreme heat. Policies designed to encourage homeowners and developers to strategically plant and preserve shade trees have potential to achieve cost-effective long-term reductions in peak electricity demand. Trees also may reduce annual electricity consumption by providing full shade over buildings during summer months; while also shedding leaves to allow warming sunlight exposure during the winter.

The magnitude of energy savings from tree shade must be understood for valid cost-benefit analysis. Although several models simulate building shade, few empirical studies estimate the actual energy savings from tree shade using household level billing data. Donovan and Butry (2009) use monthly billing data from 460 homes in Sacramento, California and find that homes with south and west facing trees have lower summertime energy bills. Pandit and Laband (2010) present a similar model for 160 homes in Auburn, Alabama and find that a 20 percent increase in tree shade reduces summertime electricity bills between 3 and 9 percent, but also substantially increase winter electricity bills. These studies identify effects based on cross-sectional comparisons of energy use between houses with varying levels of tree shade. Both models include only basic controls for home characteristics (house square footage and presence of a pool) and instead rely almost entirely on average early-spring energy to serve as a proxy of baseline energy use without shade. However, these cross-sectional estimators will be biased if tree size is correlated with other variables that affect energy efficiency, such as home age. Pandit and Laband also require participant permission to read electricity meters, introducing selection bias into estimates. Endogeneity problems are also likely if homeowners who prefer trees also practice energy conservation.

The purpose of this paper is to (i) accurately measure of the energy savings from tree shade (ii) determine whether energy savings coincide with peak electricity demand during summer months. Our quasi-experimental approach uses three unique micro-level data sources to identify the change in residential energy caused by a tree removal. We use city tree-ordinance permits to identify the location and timing of tree removals for specific households from. A twelve-year time series of monthly household electricity consumption for all households in Gainesville[1] allows for a difference-in-differences estimation using billing data from before and after a tree removal event. Data detailing structural building characteristics and used to match treatment and control homes with similar energy use characteristics.

This work makes three contributions to existing literature on the energy saving potential of urban forests. First, we use a quasi-experimental design identify a causal link between tree shade and energy use. Estimates are identified by electricity variation "with-in" households, reducing omitted variable bias and endogenity problems of cross-sectional "between" estimators. Second, our study draws from a full census of 30,000 households within the Gainesville Regional Utility service area, improving the consistency of estimates from previous studies with small samples and selection biases. Third, given the context of a city-wide tree ordinance, these estimates have direct policy relevance.

Results suggest that removing mature trees in urban setting significantly increases residential energy use, particularly during summer months. After a tree removal, households experience a 3 percent increase in average monthly utility consumption across the year. The treatment effect is largest during summer months, with an average electricity increase of 4 to 5 percent following a tree removal. Estimates suggest that energy savings from tree shade coincide seasons of peak electricity demand, providing new evidence that tree ordinances may serve as effective demand side management policies.

The first section below presents the econometric model. Section 2 summarizes the data. Section 3 presents results. Section 4 provides analysis and describes limitations of the current research design. Section 5 presents directions for future research. Section 6 concludes with an overview of important results and policy implications.

I. Empirical Model

This analysis applies an unobserved effects model[2]. In the simple case of a uniform treatment this would entail estimating a two-way fixed effects model.

$$\boldsymbol{y_{it} = \theta_0 + \sum_{t=1}^{T} \theta_s ds_t + \delta_1 aftertreat_{it} + u_{it}}$$

where $\boldsymbol{y_{it}}$ is electricity consumption in month *t* in household *i*, $\boldsymbol{\theta_0}$ is a common intercept, $\boldsymbol{ds_t}$ are month indicator variables that capture time-varying factors common to all buildings that affect and $\boldsymbol{aftertreat_{it}}$ is an indicator variable equal to 1 in periods after a tree was removed for homes where a tree was removed. $\boldsymbol{u_{it}}$ is an error term that represents unmeasured time-variant factors affecting electricity usage. The coefficient $\boldsymbol{\delta_1}$ represents the impact of the tree removal on electricity usage (i.e. the average treatment effect). The month dummy variables allow non-treated observations to enter the model to establish baseline energy use in the absence of a tree removal treatment. All variables are demeaned for estimation purposes. Additional seasonal indicator variables and interaction terms are also included to estimate seasonal treatment effects.

In this first specification, errors are clustered by household as errors are expected to be correlated within individuals, using the Stata option *robust*. A Wald test for groupwise homoscedasticity is rejected with a p-statistics of 0.0000 suggesting that heteroskedasticity is present across households[3].

Due to the persistence of seasonal temperature trends, error terms are may also be correlated between consecutive billing months. To account for this possible serial correlation, an alternative two-way fixed effects model with a first-order autoregressive error model is also estimated. The AR(1) adaptation of equation (1) can be expressed as:

$$\boldsymbol{y_{it} = \theta_0 + \sum_{t=1}^{T} \theta_s ds_t + \delta_1 aftertreat_{it} + \varepsilon_{it}}$$

$$where, \quad \boldsymbol{\varepsilon_{it} = \rho\varepsilon_{i,t-1} + u_{it}} \quad and \quad \boldsymbol{|\rho| < 1}$$

where $\boldsymbol{\varepsilon_{it}}$ is an AR(1) error term, $\boldsymbol{u_{it}}$ is a spherical error term, and $\boldsymbol{\rho}$ is an auto-correlation coefficient to be estimated using the Baltagi-Wu GLS estimator. After estimating $\boldsymbol{\rho}$**,** a transformation of the data removes the AR(1) component of the disturbance, using the *xtregar* command in Stata (Baltagi

[2] An alternative approach would be a difference-in-differences model with multiple treatment periods. However, I have not been able to find a difference-in-difference model with multiple treatment times that allows for a persistent treatment effect beyond a single period. I expect a DD model may be possible by restricting pre-post treatment periods for each unique treatment-control match, however, I am not confident that I can code/construct such a model that is theoretically correct. Any suggestions would be great!

[3] The user-written Stata post-estimation command *xttest3* uses the a Modified Wald test for groupwise heteroskedasticity in fixed effects regression models. To address heteroskedasticity the *robust* option is used.

and Wu, 1999). A Lagram-Multiplier[4] test with a null hypothesis of no first-order autocorrelation is rejected with a p-value of 0.0000, suggesting that the data does have serial correlation.

The reasoning behind possible serial correlation is that household heating and cooling systems are used differently during different times of the year. Consider a home with a broken window that is unobserved in the data. In the Spring, when neither heating or cooling is necessary, a broken window may not effect energy consumption, so the error terms for March and April (consecutive months) will be very small. However, in the summer, air conditioning will be used more intensively due to air escaping through the broken window, so error terms in July and August will both be very large.

A preferred model would include error corrections for both autocorrelation and groupwise heteroskedasticity. Stata fixed effects estimators do not support both forms of error correction. GLS estimators seem to permit both serial correlation and groupwise heteroskedasticity error corrections. However, GLS does not appear feasible given the very large sample size.

We estimate each model using a matched control group as well as the full sample of available households. Matching estimators improve the accuracy of the control baseline, especially in cases of selection bias or simultaneity bias. We use coarsened exact matching[5] to identify a single control home for each treatment home based on housing characteristics and baseline energy use patterns (Blackwell et al., 2010). Ideally, matched treatment and control homes should be "statistical twins" concerning time-varying characteristics related to energy consumption during the pre-treatment baseline. Time invariant structural characteristics correlated with time-varying characteristics are also useful matching parameters.

II. Data

Gainesville, Florida has a population of 126,000 people and a humid subtropical climate. Tree canopy covers 51 percent of the total city area, making Gainesville among the most tree dense cities in the nation. Gainesville Regional Utility (GRU) is a municipally owned electric, gas, and water utility that services more than 35,000 single family residential homes within the city. This study is restricted to single family homes within the GRU service area. Table 1 presents descriptive statistics for the treatment group, all non-treated GRU homes, and a matched control group.

[4] The Stata user-written Stata command *xtserial* uses the Wooldridge test for autocorrelation in panel data (see Drukker for details)

[5] Matching can also be done using propensity score and optimal matching (Levenshtein distance). Optimal matching may be most appropriate for matching the "life-courses" of building energy consumption.

Table 1: Baseline electricity use and home size of treatment and control groups

	Treatment				Control							
					Population				Matching			
Variables	Mean	S.D	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
winter (kWh)	1,058	777	0	6,106	1,076	779	-283	14,067	1,048	728	0	4518
spring (kWh)	842	467	1	3,870	857	519	-127	9,547	861	488	0	3,732
summer (kWh)	1,359	827	1	5,549	1,363	848	-120	13,867	1,366	837	0	5,358
fall (kWh)	1,119	610	0	4,621	1,128	623	-286	13,590	1,118	597	0	3,702
square footage	1,789	672	680	5,039	1,751	724	0	10,855	1,746	668	514	4916
N months	131	4	121	142	131	4	121	143	131	4	121	143
N homes	519				21929				519			

Variables: Average baseline monthly electricity consumption across each season (winter, spring, summer, fall) from 2001, square footage of building area, average number of monthly billing observations per home, and total number of homes. Negative electricity consumption may represent homes with solar electricity generation capacity that exceeds household electricity use.

GRU provides the full census of household level billing data for electricity, natural gas and water consumption for all months between 2000 and 2012. The dataset includes an unbalanced panel of approximately 30,000 single family residences, with a total of about 3.5 million billing records. The dataset was trimmed down to 22,448 households that had over 120 monthly electricity bills between 2000 and 2011. Unfortunately, when converting to a new data management system GRU lost about half of their billing data from the year of 2007. Despite minor issues of missing data, GRU provides micro-level utility data that identifies the household location.

The Alachua County Property Appraisal (ACPA) office maintains a rich database of housing characteristics and building permits information used for local tax records and can be easily integrated with existing electricity billing data. Due to time constraints, only a variable for home size (total square footage) is used in this analysis for developing a control group using coarsened exact matching. The ACPA dataset is described in greater detail in section V, when describing future improvements of this analysis.

Gainesville has a strictly enforced city tree ordinance that requires citizens to apply for a tree removal permit from the city arborist before cutting down any mature trees on private property. The tree ordinance also requires homeowners to replant new trees within their property following any approved tree removal. This study uses a database of 3,734 tree removal permits to identify the timing and property location of mature tree canopy loss between 2000 and 2012. The database includes the issue date, property address, number of trees and species approved for removal, and the number of trees required for replanting. Additional data about tree size and condition are available for some years. Only issue date and property location are used for the current analysis[6]. Due to difficulty merging the tree permit data based on manually entered property addresses, only 1,573 tree permits were matched to GIS property datasets, of which only 659 properties with tree removal permits fall within the GRU service area. Further restricting the sample to homes with at least 12

[6] The tree permit data is hand entered and some fields change from year to year. Working with other variables in the dataset is messy, so future data cleaning will take some time.

months of billing data before and after tree removal date 519 treatment homes remain in the final sample.

III. Results

Table 2 presents results of four unobserved effects models that differ based on treatment of the error term (robust, AR(1)) and selection of control groups (full population, matched). Model (1) is a two-way fixed effects model with robust standard errors using all households in the GRU service area. Model (2) transforms the FE model with AR(1) error for all homes. Models (3) and (4) are robust and AR(1) models, respectively, using a control groups with a single-match for each treatment home. Controls are selected using coarsened exact matching based on home square footage (10 bins), baseline 2002 summer kWh consumption (10 bins), and baseline 2002 kWh change between spring and summer (10 bins). A single control is randomly selected in cases where multiple control homes exactly match the strata of a treatment home.

Model (1) finds an average treatment effect (ATE) of a 33 kWh increase in monthly energy use when averaged across all seasons, or about a 3 percent rise in electricity consumption (base on average spring/fall electricity use of 980 kWh). An F-test of joint significance that the sum of the three[7] treatment variables is equal to zero rejects the null hypothesis with a p-value of 0.015. Total energy penalty doubles during summer months, with tree removal treatment pushing up future electricity consumption by 66 kWh per month, jointly significant with a p-value of 0.0062. The summer treatment penalty is smaller in percentage terms, with an average summer electricity rise of 4.9 percent (average summer electricity use of 1,359 kWh per month), which is about a 60% baseline-adjusted premium above the spring and fall treatment effect. During winter months, the energy penalty is negligible at 3.5 kWh per month, jointly significant with a p-value of 0.0307. Importantly, the winter treatment effect remains positive, suggesting that tree removal does not result in cold-weather energy savings by natural solar heating.

Model (2) includes and AR(1) correction for serial correlation. If the error terms for consecutive billing months are correlated (i.e. not independently distributed), then the parameters estimated in Model (1) will be biased (is direction of bias predictable?). In electricity data, serial correlation may occur due to persistent seasonal weather trends that span multiple months. Direct comparison of regression results seems to indicate an improvement in model estimation after correction for serial correlation[8]. The most notable difference between the regression results are the shifted intercepts, although this does not provide much insight into the quality of the model. The direction of average treatment effects are consistent with the results of the heteroskedastic robust model (1), however, the AR(1) has greater statistical significance for all treatment effects. Most treatment effects are smaller in model (2). The magnitude of the average treatment effect across all months drops by about 25 percent to a 24.8 kWh monthly energy increase. The average summer energy penalty drops by a third to a 44.2 kWh monthly energy increase, while the winter energy penalty doubles to 7.7 kWh per month. All relevant F-tests of joint significance are significant at the 1 percent level.

[7] The ATE across all seasons is equal to aftertreat + aftertreat*summer + aftertreat*winter.

[8] A Lagrange-Multiplier test for serial correlation using the *xtserial* command after Model(1) should formally test whether the model has first-order autocorrelation. A Breusch-Pagan test for heteroskedasticity using the *xttest2* and *xttest3* commands should also be formally tested on a model without the correction for robust standard errors. Since these regressions run for hours, I have not yet rerun models for post-testing.

Table 2. Unobserved effects model with two-way fixed effects

Dependent variable: monthly electricity usage (kWh)								
	(1)		(2)		(3)		(4)	
Variables	Population				Matching			
	FE (robust)		FE, AR(1)		FE (robust)		FE, AR(1)	
aftertreat	36.5	**	27.1	***	39.15	**	34.65	***
	(16.98)		(7.95)		(18.25)		(9.10)	
aftertreat*summer	29.6	**	17.1	**	38.84	**	21.83	**
	(13.25)		(6.78)		(15.62)		(8.64)	
aftertreat*winter	-33.0	*	-19.4	***	-26.18		-15.47	*
	(17.58)		(6.74)		(20.30)		(8.47)	
summer	243.9	***	721.0	***	-368.07	***	732.59	***
	(10.56)		(5.10)		(74.04)		(25.40)	
winter	-76.2	***	406.3	***	-655.09	***	420.24	***
	(10.37)		(4.60)		(70.84)		(23.01)	
constant	1118.9	***	620.1	***	1690.7	***	577.51	***
	(9.43)		(1.93)		(68.72)		(10.61)	
Fixed Effects								
year x month	YES		YES		YES		YES	
N observations	2,949,541		2,949,541		136,350		135,312	
N homes	22,448		22,448		1,038		1,038	
N treated obs.	27,713		27,713		27,713		27,713	
N treated homes	519		519		519		519	
R-squared	0.255				0.256			

Significance indicators: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Aftertreat is a dummy variable equal to 1 in periods after a tree was removed for homes where a tree was removed. Summer is a dummy variable equal to 1 for months June, July, and August. Winter is a dummy variable equal to 1 for months December, January, and February. Aftertreat*summer and aftertreat*winter are interaction variables capturing seasonal differences in treatment effects. Year x month dummy variables are included for all 143 billing months between 2000 and 2012. Model (1) is a two-way fixed effects model with robust standard errors using all households in the GRU service area. Model (2) transforms the FE model with AR(1) error for all homes. Models (3) and (4) are robust and AR(1) models, respectively, using a control groups with a single-match for each treatment home. Controls are selected using coarsened exact matching based on home square footage (10 bins), baseline 2002 summer kWh consumption (10 bins), and baseline 2002 kWh change between spring and summer (10 bins). A single control is randomly selected in cases where multiple control homes exactly match the strata of a treatment home.

Models using coarsened exact matching currently serve as a useful robustness check against models using information from the full sample of homes. Model (3) uses a robust estimator with one-to-one control group matching, and has higher estimates than model (1) with an ATE across all seasons of 51.8 kWh and a total summer ATE of 78 kWh per month, or an average 5.7 percent increase in summer electricity use after tree removal, both jointly significant at the 1 percent level. Model (4) includes an AR(1) error correction and has estimates more in-line with the full population estimates, with all relevant treatment effects jointly significant at the 1 percent level. The all-season ATE is an increase in monthly electricity use by 41 kWh, and the summer ATE is 56.5, representing a 4.2 percent increase in average summer electricity use following a tree removal treatment. The winter ATE using matched controls suggest an energy use increase that is larger than models (1) and (2), but still much lower than all other seasons. Interestingly, the AR(1) model appears to improve the model more under the matching sample, than in the full sample, perhaps due to additional bias from the reduced sample size.

IV. Analysis & Limitations

All models indicate that tree removal treatments increase household electricity bills across all seasons. From a policy perspective, it is useful to reverse the signs of ATEs, interpreting the *energy savings* from *tree preservation*, rather than the energy increase from tree loss. By reversing the signs of the "aftertreat" coefficients, the average treatment effect (ATE) will describe the counterfactual treatment of adding a mature tree, rather than removing a tree. In this thought experiment, results suggest an energy saving premium from tree shade during summer months, compared to energy savings during other months. We also find that trees have a minimal effect on winter electricity consumption, with no evidence of a winter heating penalty from tree shade. These seasonal differences may be explained in part by the varying intensity of solar radiation during different parts of the year. Summer months have more direct and intense sun exposure for longer periods of the day than the winter months, so winter sunlight does not heat homes as significantly. An additional factor may be that deciduous tree loose leaves before the winter permitting additional sunlight penetration to heat homes. However, an alternative explanation is that many homes are heated by natural gas rather than electricity, so a winter energy savings effect may be overlooked in these models which are restricted to electricity use.

The current identification of treatment groups introduces measurement error in multiple ways. The model assumes that all tree removal permits are identical treatments. In reality, permits are heterogeneous in the number of trees approved for removal, and the size, species and location of tree removal. In particular, some trees are not originally positioned to shade houses, so a tree removal would not identify the effect of tree shade on energy use. The timing of permit approval is a crude measure for the timing of tree removal, so any lag time between the permit date and removal data introduce additional measurement error. The current model assumes a permanent treatment effect (i.e. nothing is put in place of the removed tree), however, the tree ordinance explicitly requires new trees to be planted in the place of removed trees. Finally, the tree-ordinance only applied to large mature trees, so smaller tree removals – or illegal large tree removals – may pollute some control households. Similarly, if a removed tree shades a neighboring property, then the *defacto* treated neighbor will erroneously pollute the control group. If these measurement errors are uncorrelated with the true unobserved tree removals, then $\boldsymbol{\delta_1}$, or the ATE will be biased downwards[9] (attenuation bias) and represent a conservative estimate of the true effects. In addition, trees are more likely to be approved for removal if the tree is diseased, damaged, or dying. These types of trees may have less dense foliage than healthier trees, meaning that the treatment effect underestimates the true energy savings from more typical tree shade.

[9] However, if additional covariates are included in the regression, then the direction of bias is unknown. The summer and winter dummy variables included in the regression should not change the downward bias of the treatment effects (I think).

There is also a possibility of multiple treatment effects. If a new homeowner is more likely to remove a tree shortly after settling into a home, then the tree removal will also correspond to an occupancy change. Since occupancy is a major determinant of residential energy consumption, the multiple treatment effect will bias results. A similar problem of multiple treatments will arise if trees are removed as part of major home remodeling project or new energy efficiency upgrade. In some cases a tree may be removed when it is causing structural damage to a house (e.g. roots cracking the foundation), in which case the removal would be directly paired with a structural repair, introducing endogeneity problems.

The matched-control group is limited in this regression. Summary statistics suggest that on average, the control group is not a substantially closer match to the treatment group than the full population of households. In this case, a matched control approach may be inefficient since it restricts the sample size and does not use all available information. However, a more careful matching based on additional criterion and tighter categories (narrower bin widths) could reduce bias that may be present if the population at large is not representative of the treatment group. A more complicated technique that matches multiple control homes to each treatment home could construct a representative control group that retains maximal-information from the original sample.

V. Future Directions

We have several data sources that could help to overcome methodological problems and extend the current analysis. These datasets include additional household billing data, structural housing characteristics, details about tree removals, other permit information, and remote sensing imagery.

Data on structural housing characteristics, building permits, and other monthly utility billing data will be useful for more accurately matching control groups and mitigate problems of multiple treatment effects. The Alachua County Property Appraisal (ACPA) office maintains a rich database of housing characteristics and building permits information used for local tax records and can be easily integrated with existing electricity billing data. Records include a long list of variables relevant to energy efficiency, including; home vintage (year built), home size (square footage, construction quality, number of stories, architectural style), home occupation (owner vs. renter), home heating and cooling systems (heated square footage, AC/heating system, heating fuel type), and other physical characteristics relevant to the effect of solar radiation on energy use (building orientation, exterior wall material). These observed time-invariant structural characteristics could be very useful for constructing control groups, as they are correlated with the change in energy efficiency between different seasons of the year. For example, older energy inefficient homes are likely to have much larger fluctuations in energy use across different seasons corresponding to fluctuations in indoor-outdoor temperature differentials. We do not currently have a variable for swimming pools, which are a major source of summer energy use, but this information may be available elsewhere. Another important seasonal determinant of electricity usage is the type of fuel used for heating systems. We also possess monthly natural gas billing data over the study period that could be combined with electricity data to more accurately identify the impact of tree shade on winter energy use.

Time-variant changes in household energy consumption may also be included to account for multiple treatment effects. The ACPA data also includes information on the dates of housing sales that can be used as a proxy for changes in occupancy that may be related to tree removal. Currently, my billing data is only linked to homes, not customers. A group at the University of Florida that has agreed to provide information about changes in customer accounts across time within a house that should be a very good indicator of occupancy change. Even within a given home owner or customer account, occupancy may change with seasonal patterns that are non-random. Gainesville, Florida is a large campus-town that has seasonal occupancy patterns tied to semester schedules and also has a “snow-bird” retiree population that may leave during parts of the year. In recent years, Florida’s foreclosure crisis has also resulted in large numbers of abandoned, unoccupied housing units. We possess monthly water bills from GRU over the study time period that could serve as a proxy for major occupancy changes within a home over time. Water consumption should reflect the number of occupants more completely than electricity use or indicators of ownership change.

Another source of time-variant controls is the ACPA building permit[10] database that includes information about the date and size of home remodeling projects and energy efficiency upgrades. The UF group[11] will also provide participation data for energy efficiency upgrades through GRU’s energy efficiency rebate projects. We will take particular care to control for any new building construction, which will create a misleading correlation between tree removal and parcel-level energy use that is unrelated to building tree shade. Including energy-related time-invariant variables in the model should strengthen the causal link between tree removals and treatment effects.

Additional information about tree removals and classifications of remote sensing imagery could be used to better define tree removal treatments. Our tree removal database includes information on the number of trees removed, species of trees removed, and the required number of new trees to be replanted on the property to comply with the city ordinance. Integrating these variables with electricity data may permit testing whether effects differ by the number of trees removed, or the species of trees (e.g. evergreen vs. deciduous). For some years, the database also includes records for declined tree permits. These declined permits could be used to construct a control group; there are endogeneity concerns that the type of person who wants to remove a tree has particular energy consumption behaviors that differ from the general population. A second use of declined permits could be to assess the policy impacts of the tree ordinance, or the energy savings attributed to tree conservation. Finally, our current strategy for matching tree removal addresses with electricity data (via Google maps) has led to a relatively low successful match rate of about 50 percent. An improved matching algorithm could substantially increase the treatment sample, and may even permit a unique control group composed of households that remove trees at a future data. The detailed tree removal data still does not address the most significant source of measurement error, which is the location of the tree and the shade cast upon buildings.

Remote sensing data could be used to identify the ideal treatment variable – a continuous measure of tree shade change. High resolution aerial imagery is available[12] for 2001 and 2011. Remote sensing

[10] After a quick glance, the building permit data is somewhat messy and may require substantial work to integrate with monthly billing data.

[11] Data provided by the UF group will not include personal identification information, such as home addresses. I hope to match the datasets based upon other common fields.

[12] The University of Florida has provided 1-foot color imagery that includes information from infra-red vegetation sensors. ACPA is providing multi-return LiDAR data and 6-inch color imagery from 2001. I have also worked out a deal with a local remote sensing company to classify tree cover and tree cover change using 2001 and 2011 imagery for a reasonable fee. I then hope to use the LiDAR dataset to develop a 3-dimensional characterization of tree height.

techniques would permit a classification of existing tree canopy cover over Gainesville for both periods, and then identify the amount that tree cover changed over the 9 year period for any given property. Additional LiDAR data from 2001 may also permit height calculations for individual trees in the pre-treatment period. Given tree height and tree cover, combined with ACPA GIS data of building footprints, ArcGIS software can be used to calculate tree shade cast onto buildings. These analyses will allow a measure of tree shade loss from tree removal. Tree permits can be used to identify a more precise timing of when tree shade loss occurred. This extensive GIS data work will permit a more accurate and policy relevant answers to the question of the energy savings from tree shade. A shade calculation analysis will also permit identifying tree shade for buildings on adjacent properties, allowing for an evaluation of the negative spillovers from tree removal. The tree cover change analysis will also identify new tree cover in 2011 that did not exist in 2001. Measures of new tree growth can be combined with tree-ordinance data of tree plantings to identify the expected energy savings that can be attributed to tree replacement mandates. The promising findings of this preliminary analysis may justify the intensive GIS work needed to accurately assess the impact of tree shade on energy consumption.

VI. Conclusion

This paper contributes to literature in three ways. First, we use a quasi-experimental design to establish a clear causal relationship between tree removal and increased residential electricity use. Second, a long-time series for the census of households in the Gainesville study area eliminate sample selection issues and permit the use of matching estimation techniques. Thirdly, our results have immediate policy relevance for Gainesville, a city that is revising a long standing tree ordinance and has a municipally owned electric utility undergoing major infrastructure investments to meet a growing energy demand.

Results suggest that removing mature trees in urban setting significantly increases residential energy use, particularly during summer months. After a tree removal, households experience a 3 percent increase in average monthly utility consumption across the year. The treatment effect is largest during summer months, with an average electricity increase of 4 to 5 percent following a tree removal. Estimates suggest that energy savings from tree shade coincide seasons of peak electricity demand, providing additional evidence that tree ordinances may serve as effective demand side management policies

These are conservative estimates of the energy savings of shade trees, since part of the treatment group removed trees that were not directly shading homes. Current results are not directly comparable with previous studies because we do not use percentage tree shade metrics. Donovan and Butry (2009) and Pandit and Laband (2010) find evidence that tree shade reduces average electricity consumption during summer months, which is consistent with our findings. Pandit and Laband also find a large increase in winter electricity usage attributable to tree shade in Alabama. However, our results do not support the notion of a winter penalty from tree shade, but instead find negligible effects of tree shade on winter electricity consumption.

Future research building upon this work will help provide more direct policy prescriptions. Expanding the analysis to include tree shade metrics would establish a stronger causation between tree shade and energy use, rather than broadly defined tree removal. Further analysis of the spillover effects of tree shade, or the energy savings to adjacent households, would help to delineate the social benefits of tree cover. A targeted analysis of the effects of tree shade on peak electricity demand would add an important policy dimension to tree preservation policies. Furthermore, a nuanced analysis identifying the effects of tree placement, tree species, and types of buildings that have the largest benefits in terms of tree shade would provide practical guidance for possible demand side management policies. Finally, a counter-factual analysis to determine the benefits of tree preservation and replanting requirements of the existing Gainesville tree ordinance would help other cities assess the benefits of establishing similar policies.

References

Baltagi, B. H., Wiley (2008) *Econometric Analysis of Panel Data*.

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2010. cem: Coarsened exact matching in stata. Stata Journal. 9(4): 524-546.

Donovan, G.H., Butry, D.T. (2009). The value of shade: estimating the effect of urban trees on summertime electricity use. Energy and Buildings 14, 662-668.

Drukker, D. M. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3: 168–177.

Pandit, R., Laband, D. (2010). Energy savings from tree shade. Ecological Economics. 69, 1324-1329.

VII. APPENDIX I: TREE CANOPY CLASSIFICATIONS IN MATLAB

Motivation For Tree Cover Classification

The current identification of treatment groups introduces measurement error. The model assumes that all tree removal permits are identical treatments; however, not all tree removals change a home's tree shade equally. In particular, some trees are not originally positioned to shade houses, so a tree removal would not identify the effect of tree shade on energy use. The ideal variable of interest would be *change in tree shade* attributed to a tree removal.

Remote sensing data could be used to identify the ideal treatment variable – a continuous measure of tree shade change. High resolution aerial imagery is available[13] for 2001 and 2011. Remote sensing techniques would permit a classification of existing tree canopy cover over Gainesville for both periods, and then identify the amount that tree cover changed over the 9 year period for any given property.

The Matlab code classifies 2011 satellite imagery into three categories: (i) tree (ii) shadow and (iii) other. The method is scalable to the full Gainesville area and can be replicated using 2001 imagery to create an analysis of tree cover change for all properties. Tree permits can be used to identify a more precise timing of when tree shade loss occurred. A change analysis will also identify new tree cover in 2011 that did not exist in 2001. Measures of new tree growth can be combined with tree-ordinance data of tree plantings to identify the expected energy savings that can be attributed to tree replacement mandates.

Matlab Classification of Remote Sensing Imagery:

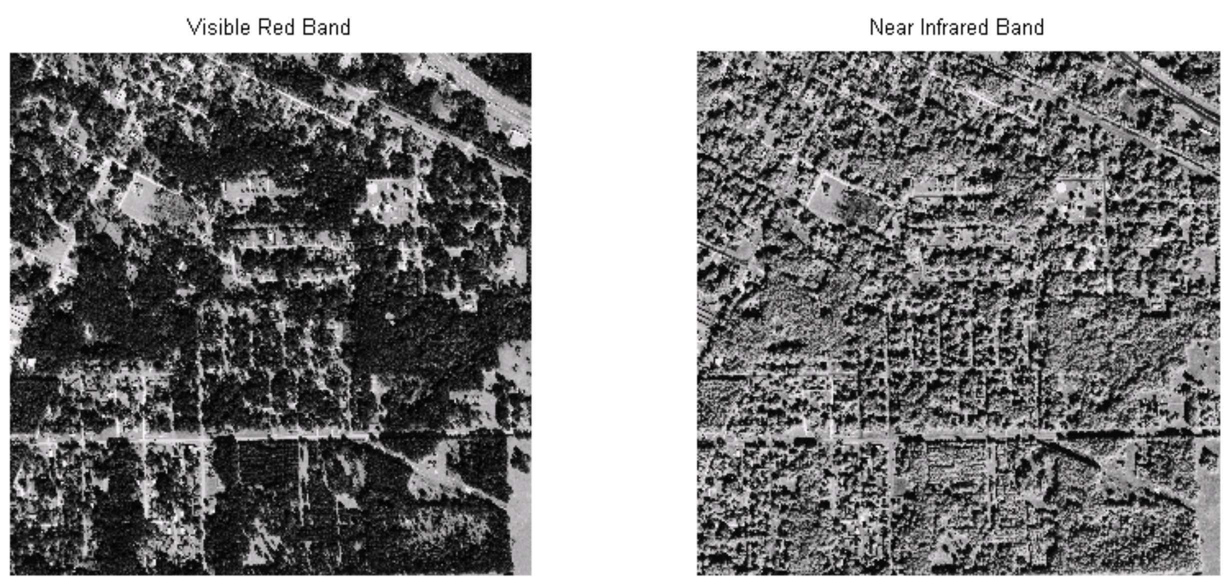
I have approximately 100 orthophoto imagery digital orthophoto tiles for 2011 that cover approximately 1 square mile with a pixel resolution of 1 foot. Each image is composed of 4 bands or channels of 8-bit image information. Each band comprises a grid of pixels containing digital numbers ranging from 0-255, and representing colors in the red, green, blue (RGB) or near infrared (NIR) portions of the electromagnetic spectrum. Band 1 is red (590-675 nm, band 2 is green (500-650 nm), band 3 is blue (400-580 nm) and band 4 is near infrared (675-850 nm). Only 3-bands can be displayed in any single image. The RGB image, displaying bands 1, 2, and 3, trees canopies are darker objects, but are difficult to distinguish from shadows. The Color Infra-Red imagery (CIR), displaying bands 4, 1, and 2 conveys additional information about tree canopies, which reflect red, infra-red, and green wavelengths, and makes a clearer distinction between trees and shadows.

[13] The University of Florida has provided 1-foot color imagery that includes information from infra-red vegetation sensors. ACPA is providing multi-return LiDAR data and 6-inch color imagery from 2001. I have also worked out a deal with a local remote sensing company to classify tree cover and tree cover change using 2001 and 2011 imagery for a reasonable fee. I then hope to use the LiDAR dataset to develop a 3-dimensional characterization of tree height.

Figures 1(a) and 1(b):

Segmenting the bands into separate matrix objects, it is clear that trees are best characterized by the red and NIR wavelengths. Figure 2a portrays the visible red band, where vegetation and shadows both show up as much darker than the surrounding environment (darker pixels are closer to zero values). In Figure 2b, the NIR image, trees are light (high values) and shadows are dark (low values).

Figures 2(a) and 2(b):



A direct ratio of NIR levels to red could be used to locate pixels containing dense vegetation (since vegetated pixels are negatively correlated). However, the result would be noisy for dark pixels with small values in both bands (shadows pixels are positively correlated). The Normalized Difference Vegetation Index (NDVI) incorporates both sources of correlation. The NDVI takes the (NIR - red) difference and normalizes it to help balance out the effects of uneven illumination such as the shadows of clouds or hills.

In other words, the NDVI is calculated on a pixel-by-pixel basis subtracting the value of the red band from the value of the NIR band and divide by their sum. The backslash operator in Matlab can calculate NDVI very efficiently for large matrices, single line of code: ndvi = (NIR - red) ./ (NIR + red).

The NDVI falls on a scale of -1 to 1. Generally, larger NDVI values represent high chlorophyll content, with a threshold somewhere between NDVI>0.1 and NDVI>0.4 characterizing tree canopy. Within this range the proper threshold can vary from image to image depending on the time of day the image was taken and the amount of shadows in the image. The threshold is best identified by visual inspection. Graphing the NDVI on a color scale in Figure 3(a), permits visual comparison against tree cover in the CIR imagery. I manually identify an NDVI threshold of 0.25 as a lower bound for characterizing tree cover (NDVI>0.25=tree). Since I am also interested also interested in shadows (i.e. tree shade on houses), I also use the NIR band (Figure 3b) to set an upper bound threshold of 0.1 to characterize shade (NIR<0.05=shaddow).

Figures 3(a) and 3(b):

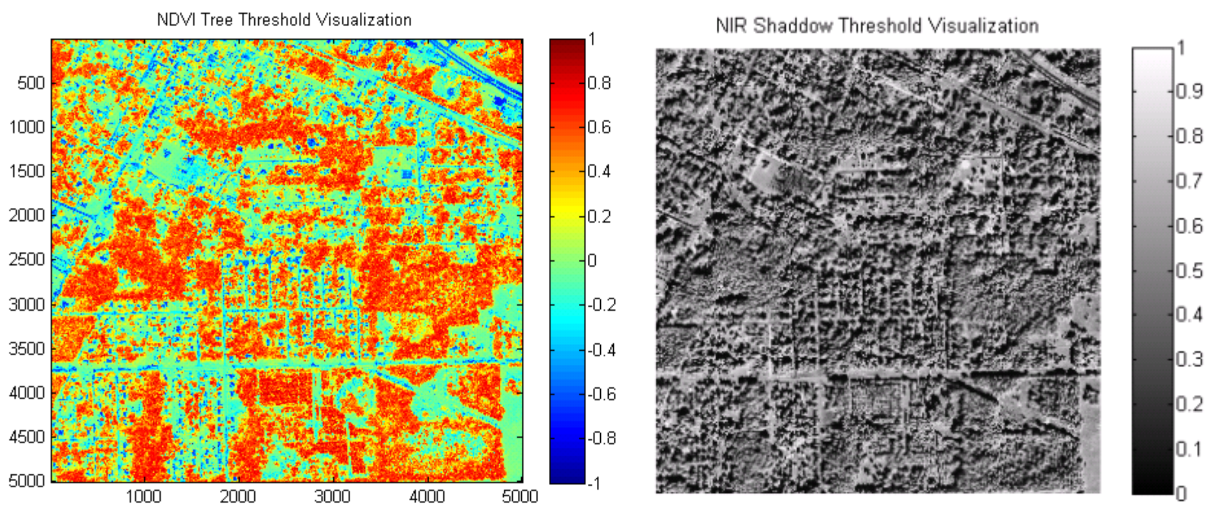
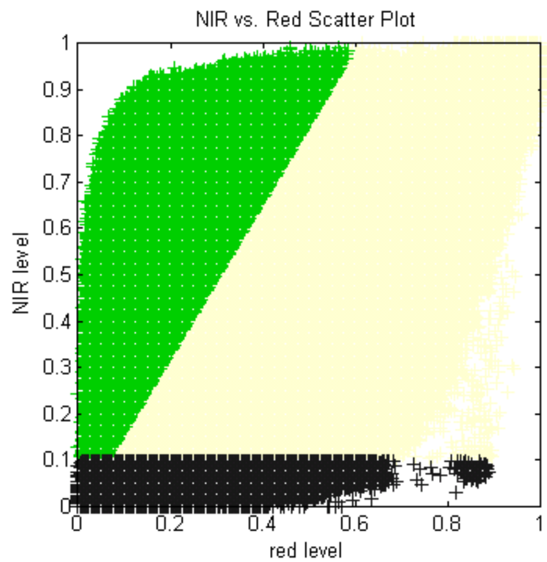
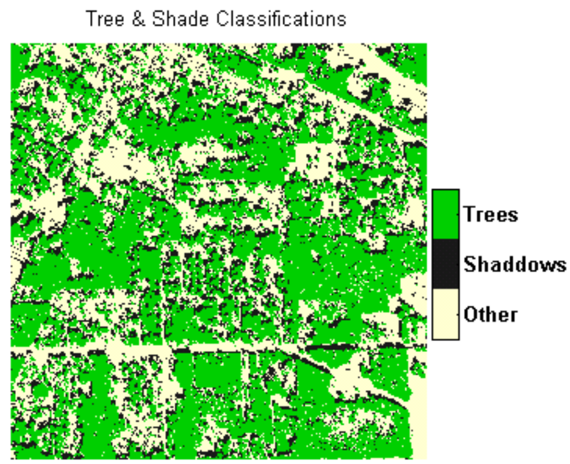


Figure 4a visually portrays the thresholds on a scatter plot of correlations between the red and NIR bands, where each (+) represents a single pixel in red-NIR space. The Northwest area of the graph, where the NIR level is well above the red level, are pixels characterizing vegetation (green +). Shadows represent the bottom portion of the NIR band (black). Figure 4b represents the tree and shaddow classfication map when pixels are reclassified to three categories in a new image (or matrix).

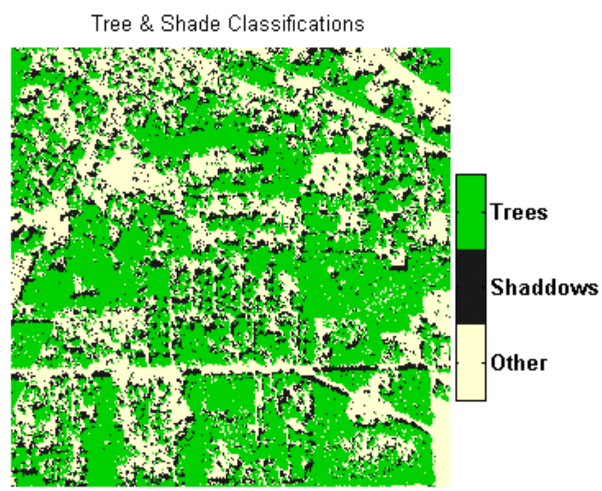
Figures 4(a) and 4(b):



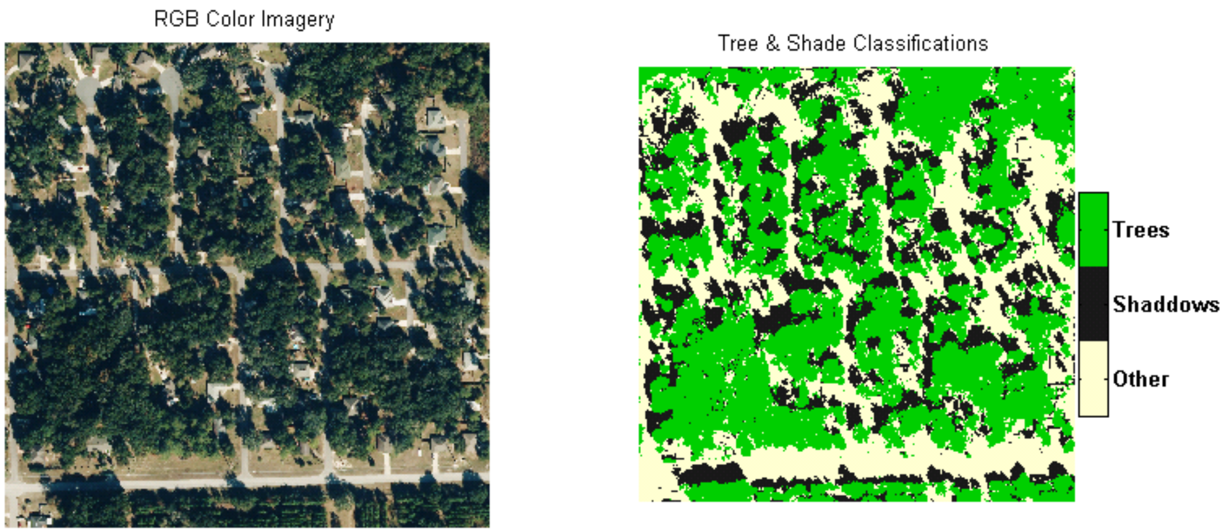


Upon visual comparison, these tree and shade classifications appear remarkably accurate, at both the scale of the full orthoimage (~1 square mile), as in Figures 5a and 5b, and at zoomed in neighborhood scales, such as Figures 5c and 5d.

Figures 5(a) and 5(b): Tree and Shade Classifications



Figures 5(c) and 5(d): Zoomed Classifications



Future work will involve (i) automating this process over all 100 images (a simple for loop), (ii) replicating the analysis on earlier imagery from 2001, and (iii) taking the difference of tree cover from the two years to identify changes in tree cover and tree shade. GIS shapefiles of building footprints for all households can be used to identify the amount of each building shaded before and after a tree removal. The tree permits are still useful for providing a clean timing of when the change in tree shade occurred.

The results presented in this study are conservative estimates of the energy savings of shade trees, since part of the treatment group removed trees that were not directly shading homes. This appendix demonstrates a scalable method for identifying tree cover and tree shade using publically available satellite imagery. Expanding the analysis to include tree shade metrics would establish a stronger causation between tree shade and energy use, rather than broadly defined tree removal. Further analysis of the spillover effects of tree shade, or the energy savings to adjacent households, would help to delineate the social benefits of tree cover.