THESIS

A COUNTY-LEVEL ANALYSIS OF RESIDENTIAL SOLAR ADOPTION IN THE UNITED STATES

Submitted by

Matthew Kerschen

Department of Economics

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Spring 2012

Master's Committee:

Advsior: Robert Kling
Co-Advisor: Sammy Zahran

Craig Bond

ABSTRACT

A COUNTY-LEVEL ANALYSIS OF RESIDENTIAL SOLAR ADOPTION IN THE UNITED STATES

This thesis set out to achieve two major objectives, with a third objective added at the end. The first was the update and analysis of the zero-inflated negative binomial (ZINB) model used by Zahran et al. (2008) in regards to its validity and robustness as a predictor of the count of solar using households in a county. The second objective was to use the model to provide an empirical measure of the effect financial and regulatory incentives have on the count of solar using households. The final objective was to explore and explain an unexpected decrease in the count of solar using households. This was done by using a ZINB regression to model the number of occupied housing units that use solar heating at the county level over the period from 2000 to 2009. In addition to analyzing the effects of the explanatory variables, geographic information systems (GIS) modeling was used to provide geographic mapping of the distribution of occupied housing units that use solar heating. The results indicate that Zahran et al.'s (2008) model is a robust and accurate predictor of the count of solar using households. Financial incentives were found to have an insignificant impact on the count of solar-using households, while regulatory incentives decreased the odds of a zero count in a county, but also decreased the expected count. A correlation was found between densely populated counties and the decrease in the count of solar using households.

TABLE OF CONTENTS

Chapter 1: Introduction

With the increased awareness of the true costs of fossil fuels and the growing push for protecting the environment, more people and governments are turning towards renewable energy sources such as solar energy. Most forms of renewable energy are still not economically feasible when compared to more traditional sources of electricity. In order to address this issue, governments offer a variety of financial incentives such as subsidies and loans as well as enforce various policies such as requiring that a certain percentage of a utility's energy comes from renewable sources. One of the main aims of this thesis is to estimate the effectiveness of financial incentives and regulations enacted between 2000 and 2009 that were designed to encourage residential solar energy use.

There are several benefits to solar energy that have caused governments to support the growth of the solar panel industry. Provided the location is good, the simplest reason is that solar panels are a reliable renewable energy source. The other reasons have to do less to do with how much energy is provided and more to do with the way in which the energy is provided. Solar panels receive their energy from the sun and thus only produce energy when the sun is out. This mostly corresponds with peak energy consumption times, when the demand for energy is greatest. This means that solar panels provide the most energy at the times when energy is most needed. This can be very beneficial to utility companies because normally they have to build and operate extra plants to deal with these peaks in energy demand. The other benefit of solar panels is that they can be installed on private homes and businesses. On the individual level this allows households and businesses to produce a significant portion of their energy themselves, and with expected improvements in the performance of solar panels, solar energy producing households may one day achieve self-sufficiency. On a societal level this is a step away from a centralized model of energy supply and toward a more distributed model where energy can be supplied at or close to the point of use. This reduces the need for transmission facilities and land dedicated to energy production.

It is necessary to understand what motivates consumers to purchase solar panels for their homes so that the limited support governments can provide is directed in the manner that would bring the greatest possible benefit. There are many factors that could influence a consumer's decision, both financial and nonfinancial in nature. They include how much the consumer values the environment, how energy conscious the consumer is, how long they plan to live in their house, and so on (Palm; Faiers). Of course there is little governments can do to address those issues aside from running advertising/information campaigns promoting the environmental benefits of solar energy use. What governments are interested in is the impact financial incentives and regulations have in promoting solar system adoption by consumers.

This thesis then will address how basic climate, economic, and sociopolitical factors influence a consumer's decision as well as the impact that financial incentives and other governmental policies have on the adoption of solar heating. This will be accomplished by extending the work of Zahran et al. (2008) by: 1) updating their statistical model with new data from 2005-2009; and 2) including measures of financial incentives and regulations enacted between 2000 and 2009 to empirically assess the effectiveness of government efforts to encourage solar energy adoption. It is important to note that this thesis, as Zahran et al. did, measures solar heating units specifically and not solar electricity systems such as photovoltaic (PV) panels due to the more comprehensive data available on solar heating adoption compared to PV adoption. These two technologies are very similar, however, and a consumer deciding whether to install a solar heating system has much the same motivations and issues as he would face if deciding to install a PV system instead. Therefore solar heating will be acting as a representation of the entire solar industry in this thesis.

There was an unexpected trend discovered while analyzing the data that also needs to be addressed. That trend is that residential solar usage has decreased by a significant amount between 2000 and 2009 despite an overall increase in the number of households during that time. Combined with the rising price of fossil fuels and drop in the costs of solar technology, the fact that instead of an increase in solar users there was a decrease is rather difficult to explain. An interesting correlation was found between a decrease in solar users and an increase in population density in highly urbanized counties that could explain it, however. This thesis will first explore this correlation before explaining how this could be the cause of the trend and some of the implications and ramifications if so.

Chapter 2: Literature Review

The literature on consumer adoption of solar systems can be divided into three broad categories; 1) technical papers dealing with a specific aspect of the technology, method of installment, or feasibility; 2) policy papers that analyze the effect of government policies; and 3) social-psychological papers that focus on consumer motivations to purchase solar systems.

Technical Papers

A number of reports focus on the technical aspects of solar energy adoption such as the best type of system for an area or the best angle for the panels. For example Payne (2000) explores the feasibility of a new type of PV technology. His focus is on how much more efficient the technology is and how much cost-savings it would generate. Holbert (2007) instead focuses on one area, Phoenix Arizona, and calculates the best direction and angle for solar panels to generate the most energy. As he is looking at just one town he is also able to discuss the specific regulations and incentives that affect the local solar market. Perez (2004) also focuses on the technical side of the discussion; although in his case he is looking into financial technicalities such as the proper way to measure profitability of solar systems for homeowners.

Financial and Regulatory Papers

Just as there have been a variety of government policies and incentives aimed at promoting residential solar panels over the last 40 years (both in the U.S. and in other countries around the world), there have been a number of studies done that attempt to determine how successful these policies and incentives were and what made them succeed or fail. Some studies focus on analyzing specific government programs, for example Hoffman and Kiefer's (2001) study on the German 1000 rooftop program or Haas' (1998) study on the Austrian rooftop program. Still other studies focus on a specific locality and try to determine the best policies for it, given local conditions. One such study is Holbert's (2007) study of the Phoenix area. Other studies apply a variety of methodologies in order to answer the question, such as Long's (1993) econometric analysis, Hasset's (1993) use of panel data, and Hayne's (2002) use of case studies.

While the methodologies and specific focus of the studies are all very different, there are some overarching ideas and conclusions made about the residential solar panel industry. The first is that, in general, financial incentives work for developing the industry by lowering the costs of solar panels. This is because as the industry develops, suppliers and installers get better at making and installing solar panels. As a result, they reduce costs and can therefore charge customers less. Also as the market grows, suppliers are able to build bigger factories and take advantage of economies of scale. Where the studies differ is in what financial incentives work the best at doing this.

Another common conclusion is that without the necessary infrastructure in place, financial incentives will have a very limited impact. In other words, if there are not enough installers or quality control checks, or if it is difficult to actually connect the solar panel system to the energy grid, then it will take a long time to completely install all the new solar panel systems that the financial incentives will bring. This will hurt the reputation of the program and the industry, leading to fewer consumers buying solar panels.

Similarly to the previous conclusion, many studies such as Haynes (2002) and Painuly (2001) have found that quality control is integral in both lowering costs and ensuring the reputation and thus success of a program or incentive. Programs or incentives that have a poor quality control process not only lead to more solar panel failures, thus damaging the reputation, but also fail to properly promote industry growth. This is caused by the fact that poor quality control means that installers will be sub-par and unlikely to improve and will also lead to poorly designed solar panels being supported on the market by the incentives, lowering the overall quality of the industry.

Haynes (2002) also states that a stable and adequate source of funding is vital for any financial incentives. This is not quite as crucial at the residential level as it is at the utility level, where many incentives and contracts are very long-term, but it still plays a significant role in the residential market as well. For example, if funding runs out in a given year, or there is not a guarantee there will be enough funding, then consumers are more likely to wait and not purchase a solar panel system until they can be sure that they will receive the funds from the incentive.

Socio-Economic Papers

There are a number of papers, especially in the last ten years, which look into what motivates consumers to purchase solar systems and go beyond just the financial aspects of that decision.

Palm (2011) conducted a series of interviews with consumers in Sweden who had adopted green technology, either PV or microwind turbines, or were in the process of doing so, to uncover their motivations and any barriers to adoption. There were two major motivations reported for adopting green technology. One is for environmental reasons, both in the strict sense of wanting to reduce pollution and also as being an integral part of their self-image of living a green lifestyle. The other motivation is to be more independent, whether that is the desire to be more independent from the electricity companies, society, or to be more financially independent, the underlying motivation is that green technology makes them less dependent on others. The major barrier to adoption is the high upfront cost combined with the long pay-back period.

Faiers (2009) conducted an in-depth review of residential solar adoption in the UK in order to determine why several incentive programs were succeeding while another program was failing. Faiers analyzed differences in motivations of early adopters as compared to early majority adopters, the next consumer group to adopt a technology. What he finds is that early adopters base their decision primarily off of environmental concerns and their interest in solar technology. While those motivations are still important to the early majority, they also care about aesthetic, financial, and operational issues and find solar technology lacking in those regards.

Gillingham (2010) and Rothfield (2010) both look at the impact that previous adoption of solar technology in an area has on future adoptions in that area. They both used different methods: Gillingham ran three different experiments each analyzing different levels of data and Rothfield used a zero-inflated negative binomial model that allowed her to also analyze the effect of several other variables. In the end they both came to the same conclusion that a previous adoption in an area (both at the street level and the zip-code level) increases the likelihood of further adoptions in that area. The two most likely reasons for this are that seeing a neighbor with a solar system motivates people to adopt solar as well in order to not be outdone, or that they will learn more about solar systems from the neighbor and with this increased knowledge on the subject, feel confident enough to adopt the technology as well.

Benchmark Study Review

Finally the article "Greening Local Energy" by Zahran et al. (2008) endeavors to help policy makers design appropriate policies for household adoption of solar energy by looking at county-level data for the entire country. Zahran et al. do not compare the different policy tools available, but instead provide all the other information a policy maker might need, such as the location of households that are already using solar energy as well as pertinent environmental, economic and sociopolitical factors that would explain why households are willing to pay to install solar panels on their house.

Zahran et al. maintain that the most important factor to consider is how much solar radiation a county receives, as higher amounts of solar radiation produce more solar energy per square foot, requiring fewer panels to collect the same amount of energy. This means that solar energy systems will be cheaper and take up less space, both of which should encourage adoption. Another important environmental factor is the climate of a county, which they measure using maximum temperature. This is important as counties that have a hot climate do not have as great a need for heating while counties with a cold climate are much more likely to have their solar panels damaged in winter periods.

Solar energy systems have very high up-front costs so naturally economic factors play a large role in solar energy adoption. The most obvious factor to consider is how wealthy somebody is, as the wealthier they are, the less of a financial burden it is to install a solar energy system. The article uses the median home value of a county to measure this, as this gives a general idea for how wealthy the county is while also indicating a source of ready capital to finance installation. Just because somebody is wealthy does not mean that they are willing to actually spend that money on expensive goods. Many studies, such as Gourinchas and Parker (2002), have shown that consumers between the age of 40 to 49 are the most likely to purchase expensive durable goods, such as solar energy systems, and so Zahran et al. also measure the percentage of the population in a county between age 40 and 49. The article also measures whether there is a solar energy retailer in the county or not. Without a retailer nearby, adoption is unlikely. Another economic factor measured by Zahran et al. is the level of urbanization in a county. This is an important factor largely due to the fact that more urbanized counties have dense social networks that increase the effectiveness of word of mouth propagation of the technology as well as informational campaigns.

The last set of factors that predict solar energy adoption are the sociopolitical factors. In short there are major environmental benefits to using solar systems and so counties with populations that are both aware of and care about this are more likely to adopt solar systems. The article uses three different measures to determine the likely attitude of a county. The first is what percentage of the population are Democrats, as Zahran, Brody, Grover, and Vedlitz (2006) found that Democrats are more likely to adopt policies designed to conserve the environment and thus more likely to switch to solar energy. The next is the number of nonprofit environmental organizations in the county. Environmental nonprofits function to increase awareness of environmental issues, including the benefits of solar energy. The last measure is whether or not a county is a member of The International Council for Local Environmental Initiatives (ICLEI).

ICLEI works to help members promote environmental sustainability, which means that the county is invested in working on those issues so members of the community are more likely to adopt solar energy.

After explaining the reasoning behind all the factors the article constructs a statistical model to determine how many households have adopted solar energy for home heating. The technical problem they had in modeling the count of solar energy users in a county was that many counties had zero households that use solar energy, which invalidated most ordinary statistical approaches. A common way to account for excess zeros is to use a zero-inflated count model as advanced by Lambert (1992). In the end they used a zero-inflated negative binomial regression (ZINB regression). The ZINB regression splits the calculation into two smaller models, one that simply determines whether or not a county will have any households that use solar energy and another that determines just how many households use solar energy in counties that do have some household users.

They created four separate models by starting without using any of the factors and then adding in first environmental, then economic, and finally sociopolitical factors. In the end the final model that incorporates all of the factors was the best predictor of household adoption. Overall their statistical model did a good job of predicting household adoption although it did systematically overestimate zero count counties.

Next they use predicted values to create standardized residuals to determine which counties were significantly above or below expectations from the model. If the model is a good predictor for solar energy adoption based on environmental, economic, and sociopolitical reasons, that means counties that are performing above expectations are likely providing their households with additional incentives such as financial assistance and other policies while counties that are performing below expectations have untapped potential and that proper policies could provide the boost needed to get households started.

Zahran et al. end by discussing the significance of spatial predictors. Solar radiation is extremely significant with counties having high levels of solar radiation being very likely to have households using solar energy. Maximum temperature was also a significant predictor. Using the square of maximum temperature, Zahran et al. were able to show that as expected, counties with either high or low temperatures were less likely to have households using solar energy. On economic factors, measures of wealth, urbanization, and the percent of the population aged 40 to 49 were all significant positive predictors of household adoption. The presence of a solar energy retailer was not significant however. This is likely due to using a coarse measure, however the non-significance of retailer presence as a predictor of the count of solar energy users in a county, might also mean that it is easy for households to purchase their solar energy systems from outside the county and online. For the sociopolitical factors, the percentage of the population that vote Democrat and whether the county is a member in ICLEI are both significant positive predictors and have at least as great an influence on solar energy adoption as solar radiation received. The number of environmental nonprofits in a county did not have any impact however and no explanation is given as to reasons behind this.

The article shows that environmental, economic, and sociopolitical factors play an important role in determining household adoption of solar energy. Zahran et al.'s paper serves as a basis for policy makers to design their policies around. Moreover, the inventory of variables analyzed serve as a basis for the investigation of the effect of financial and regulatory incentives enacted between 2000 and 2009 on present period solar adoption.

Chapter 3: Data Set and Variables

This chapter describes the data collection efforts and measurement decisions. To enable valid and reliable comparison, where possible the same variables used in Zahran et al. are used in this thesis.

Dependent Variable

To measure household adoption of solar technologies the variable *occupied housing units with solar heating* is used. While the data source for this variable has changed from the long form of the 2000 Census to the 2005-2009 American Community Survey (ACS), the question asked remains the same and the ACS provides a comparable sample to the long form Census. The questions asked is this: "Which FUEL is used MOST for heating this house, apartment, or mobile home?"

Baseline Variable

The number of *occupied housing units* in a county is used as a baseline variable as the more housing units there are the more chances that one of them will have a solar heating system. The data come from the 2000 census and the 2005-2009 ACS.

Environmental Variables

The viability and effectiveness of solar heating is dependent on how much solar radiation an area receives and the climate of the area. Since these are both long-term variables that change little over a 10 year period, the data from Zahran et al. was used. Therefore the data for *solar radiation* is the same used in "Greening Local Energy". That data originally came from the National Renewable Energy Laboratory (NREL), where they modeled the average total solar radiation in $kWh/m^2/day$ for $40km^2$ grid cells.

Zahran et al. then took that data and used GIS to fit the grid cells to specific counties. They then took the weighted average of all the grid cells in each county to arrive at an average solar radiation for each county as a whole.

The climate variables were replaced instead of simply updated with new data. Zahran (2008) used the maximum temperature and maximum temperature squared for a county in 2000. In this thesis, *mean maximum temperature*, which is the mean temperature in July for the climate period 1941-1970, and *mean maximum temperature squared* were used. *Mean minimum temperature*, which is the mean temperature in January in the climate period 1941-1970, is also added. This variable was added only after testing to ensure that there was no multi-collinearity with *mean maximum temperature*. *Mean maximum temperature squared* is used to determine if there is an effective temperature range outside of which solar adoption rapidly decreased. The new variables come from the Area Resource File (ARF) and are measured in degrees Fahrenheit. This older climate data does a better job of representing the actual climate of a county and not just what the weather was like in a given year.

Economic Variables

The original report measured four economic variables: median home value, solar energy service providers, urbanization, and percentage of population aged 40 to 49. The *median home value* serves as an indicator of the amount of available capital the households have with a higher home value increasing the likelihood of adopting solar heating. The data for the updated variable comes from the 2005-2009 ACS, again measuring median home value by the value estimated by the occupants, with the prices adjusted for inflation using the CPI indexed to 2000.

The data for *solar energy providers* is carried over from the original report without being updated both because the original report found it to be statistically insignificant and because of the difficulty of collecting reliable and valid information. These data originally came from the Solar Energy Industries Association (SEIA) and measures whether or not the SEIA reported the presence of at least one solar energy provider in the county.

It is impossible to update the *urbanization* variable until 2012 or 2013 and so a replacement variable was needed. Urbanization was used as a proxy to measure the presence and strength of a community for word of mouth to spread through. *Population density* was chosen as the replacement and is measured by the number of people per square mile of land area in a county. The ACS does not measure this and so the 2010 Census was used instead, meaning that the values come from 2010 instead of 2005-2009 and so do not perfectly match the time logic of other variables used.

Finally there is the *age 40 to 49* or *consumption age* variable that is used based on the fact that people of that age are the most likely to adopt expensive durable goods, like a solar heating system. This measures the percentage of the population that is between ages 40 to 49 in a county. Again the data for this variable is updated using the 2005-2009 ACS.

Sociopolitical Variables

The three sociopolitical variables used in the original report were the net Democrats in a county, the number of environmental nonprofit organizations, and whether the county is a member in the International Council for Local Environmental Initiatives (ICLEI).

Net Democrats is measured because Democrats are generally much more environmentally conscious and thus more likely to both adopt solar technologies and support policies and regulations that encourage solar adoption. The data for the variable has been updated from measuring voting in the 2000 presidential election to the 2008 presidential election where the percentage of voters for McCain was subtracted from the percentage of voters for Obama to provide the net percentage of Democratic voters. The data comes from http://uselectionatlas.org/.

The number of *nonprofit environmental organizations* in a county ended up being statistically insignificant in the original report and so the data was not updated[1]. The data originally came from the National Center for Charitable Statistics (NCCS).

Membership in *ICLEI* serves as an indicator of willingness and likelihood of the local governments to adopt policies and regulations that promote environmentally positive behavior, such as adopting solar technologies. Unfortunately the original data on membership in the ICLEI in 2000 was lost and so both periods use the same data from 2008. This data is drawn from the ICLEI website (www.iclei.org).

New Variables

Several variables have been added to the model in an attempt to better estimate the adoption of solar technology, and to pursue the main novel objective of this thesis: estimating the effect of regulatory and financial incentives enacted between measurement periods on household adoption of solar energy.

[1] An attempt to update this variable was made using only the data available for free from the NCCS, but this updated variable produced similar results to the original variable and so wasn't included in the model.

The first variable to be added was *percentage of population that has lived in their current home for at least five years*. This was meant to account for the fact that solar technology has a long payback period and so is more likely to be adopted in areas where people stay in the same house for a long time. It was also meant to measure and account for the increase in social/geographic mobility of the country. Unfortunately the ACS data is not comparable to the Census data because it is a multi-year estimate. This means that somebody surveyed in 2005 might, at that point in time, have only lived in their house for 4 years and thus not count towards the variable while if they had been surveyed in 2007 they would count. Since the new data for the variable cannot be compared with old census data the variable had to be dropped.

Next *median household income* was added to better account for a household's ability to afford a solar heating system. The 2000 data for this variable comes from the 2000 Census while the 2005-2009 data comes from the ACS. The CPI indexed to 2000 is used to account for inflation. This variable was added only after testing to ensure that there was no multi-collinearity with *median home value*.

A variable to control for time was also added as the dataset contains both the 2000 values for all the variables as well as the updated 2005-2009 values. This means that there are two observations for each county, one for 2000 and one for 2005-2009. Therefore *pre/post* is used as a dummy variable to indicate whether an observation is from 2000 or 2005-2009.

Finally a set of financial incentive and solar regulation variables were added. *Total financial incentives* and *total solar regulations* variables were created to indicate the total number of incentives or regulations present between 2000 and 2009 relevant to solar heating. These data came from the DSIRE website (see appendix 1 for more details). In order to more accurately estimate the effect of these incentives and regulations difference in differences was used to account for heterogeneity, selection bias and time. Difference in differences is an econometric technique where the dependent variables of groups that have implemented the treatment, in this case the incentives and regulations, are compared to the dependent variables of groups that have not implemented the treatment, essentially control groups, both before and after the treatment. The difference between the groups before the treatment is then subtracted from the difference between the groups after the treatment to arrive at the difference in the differences, or the actual impact of the treatment by accounting for time and structural effects. That at least is how it is done in the simple case; in this more complex situation the difference estimator was instead worked into the regression, however the underlying theory and assumptions remain the same. *Financial difference estimators* and *regulation difference estimators* were created by multiplying *Total financial incentives* and *total solar regulations* by the *pre/post* variable.

As the updated solar count as well as a number of other variables comes from the 2005-2009 ACS 5-year estimate, the new period will be referred to as 2005-2009 to better represent the fact that much of the data was gathered piece-meal over that time-period. Despite being a 5 year estimate instead of a sample collected all at once as the 2000 Census was, most of the variables (and all those actually used in the model) can be meaningfully compared between the two.

Chapter 4: Statistical Procedure

As this thesis builds on Zahran (2008), the first step was to determine if Zahran's model was still robust and an accurate predictor of the dependent variable with the new data. Zahran used a zero-inflated negative binomial regression (ZINB) to model the data. ZINB regression works by estimating two separate models at the same time. The first is a logit model that gives the conditional expectation of the probability that there will be zero households with solar heating:

$$\psi_i = \Pr(A_i = 1 | \, z_i) = \frac{\exp(\gamma z_i)}{1 + \exp(\gamma z_i)},$$

Where A = 1 indicates membership in the always zero group so this is measuring the probability of an observation to fall in that group conditioned on the inflation variables z. The second is a negative binomial model that derives the conditional expected number of households with solar heating:

$$\Pr(y_i \,| x_i, \;\; A_i = 0) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(x_i \beta)} \right)^{\alpha^{-1}} \left(\frac{\exp(x_i \beta)}{\alpha^{-1} + \exp(x_i \beta)} \right)^{y_i}.$$

This is the probability of an observed count of y conditioned on both the x-variables and that the observation is not part of the always zero group. Β represents the coefficients of the x-variables. Γ is the gamma function and α represents unobserved heterogeneity and determines the degree of dispersion (Long, 2006). Expanding on that, α measures how much the variance diverges from the mean. The gamma function is well known and is an extension of the factorial function that works with real and complex numbers. In this case it is used to generate a gamma distribution of the error terms introduced with α. This gamma distribution when combined with the Poisson distribution forms the negative binomial distribution and allows the negative binomial regression to relax the assumption that the variance equals the mean.

The first step to validate the use of this model is to determine if there is still an excess of zero observations. It turns out that there are even more counties with no households using solar heating, up from 1,736 in 2000 to 2,126 in the 2005-2009 measurement period. Therefore the use of a zero-inflated regression is a valid method to ensure that the excess zeros do not skew estimates. There is another zero-inflated regression besides the ZINB, the zero-inflated Poisson regression, however, and as that is the more commonly used regression, there is still the need to determine that the ZINB is a better fit for the data. This is done by comparing the variance of the dependent variable (*nsolar*) to the mean, as the Poisson regression assumes that the mean and variance are equal, while the negative binomial regression does not. The results were that the updated count of *nsolar* had a variance of 2,876 and a mean of 10. This is not as large of a mismatch as Zahran et al. found, with the original *nsolar* having a variance of 15,669 and a mean of 13, but it is still a large enough disparity to invalidate the assumption underlying the Poisson regression. These findings show that ZINB regression is still a valid method of modeling the data and will be the method used.

Chapter 5: Results

Summary Statistics

To begin, the summary statistics of the variables for both 2000 and 2005-2009 are reviewed in order to highlight any significant changes in the key characteristics of the population. Table 2 below includes the summary statistics of all the variables, while Table 1 has the total number of households that rely on solar heating in each period. The most obvious change comes from Table 1, where the total number of occupied housing units that rely on solar heating has decreased by almost a quarter despite a substantial increase in the number of occupied housing units.

This is especially perplexing as, theoretically and in Zahran's report, all of the variables in Table 2 that have a positive coefficient increased and thus an increase in solar households would be expected. It is obvious that there is some hidden shift or trend in the data that the summary statistics don't reveal that explains this unexpected result. Several different methods to uncover the underlying reason were attempted before a possible explanation was uncovered. As those previous attempts help illustrate what has and hasn't changed from Zahran et al.'s report they are still presented below.

Table 1. Total number of occupied housing units with solar heating by year

Solar count	
Period	Total
2000	40940
2005-2009	31126
Occupied Housing Units (thousands)	
Period	Total
2000	104852.3
2005-2009	111938.3

Table 2. Summary Statistics

2000					
Variable	Obs	Mean	Std. Dev.	Min	Max
Solar count	3108	13.17246	125.1359	0	6349
Occupied Housing Units[#]	3108	33.73627	104.6246	.031	3133.774
Solar Radiation	3108	4.339217	.7915084	2.84	7.496875
Temp Min	3111	32.91035	12.02736	1.1	67.2
Temp Max	3111	75.85616	5.353499	55.5	93.7
Temp Max Squared	3111	5782.807	799.391	3080.25	8779.69
Population density	3109	244.9088	1675.725	.1	66940.1
Home Value[#]	3107	84.19517	47.37867	13.8	1000.001
House Income[#]	3108	35.26892	8.837691	12.692	82.929
Consumption Age	3108	.1500682	.0146798	.0795966	.2835821
Net Democrat	3108	-.215297	.2505282	-.848	.799
ICLEI	3108	.0357143	.1856067	0	1
2005-2009					
Variable	Obs	Mean	Std. Dev.	Min	Max
Solar count	3109	10.01158	53.62467	0	2188
Occupied Housing Units[#]	3109	36.00459	108.9863	.041	3178.266
Solar Radiation	3108	4.339217	.7915084	2.84	7.496875
Temp Min	3111	32.91035	12.02736	1.1	67.2
Temp Max	3111	75.85616	5.353499	55.5	93.7
Temp Max Squared	3111	5782.807	799.391	3080.25	8779.69
Population density	3109	261.4817	1733.225	.1	69468.4
Home Value[#]	3109	106.738	73.43501	24.41757	830.5305
House Income[#]	3109	35.9616	9.484394	15.67126	94.10981
Consumption Age	3109	.1460875	.0171752	.0464611	.245614
Net Democrat	3109	-.1531736	.2760695	-.8773	.8592
ICLEI	3108	.0357143	.1856067	0	1
Financial Incentives	3113	4.699647	3.601368	0	16
Solar Regulations	3112	5.946996	3.569646	0	13
#=measured in thousands		Inc. and Reg. not measured in 2000			

One way to shed more light on the drop in occupied housing units that rely on solar heating is to look into the geographic distribution of solar using housing units and how that distribution has changed between the periods. The best way to do that is through maps, dividing the count of solar energy users in a county into quartiles. Figure 1 shows this for the 2000 data while Figure 2 shows this for the 2005-2009 data. From looking at these maps the only systematic change seems to be the large increase in counties with zero solar-heated occupied housing units stretching from the Carolinas to Texas. To focus more directly on changes rather than levels, Figure 3 shows this by splitting the counties into those that showed a significant drop in occupied housing units that rely on solar heating, those that had little change, and those that showed a significant increase in occupied housing units that rely on solar heating between the two periods.

Figure 3 tells a rather different story from that suggested by simply eyeballing the changes between Figure 1 and Figure 2. Instead of the single region suffering disproportionate losses, as Figure 2 seemed to show, there is instead no readily apparent pattern to the losses, as there is almost always a county that grew right next to a county that suffered losses and both are interspersed with counties with no significant change. Another fact that Figure 3 reveals, with reference to the legend, is that one county suffered a 4,161 decrease in occupied housing units that rely on solar heating, which accounts for almost half the overall decrease observed between the two time periods.

This county is Los Angeles County, California. The specific reasons for such a large drop in the county with the highest number of occupied housing units that rely on solar heating are unclear, as none of the explanatory variables can account for such a large decrease. Therefore a new test is needed in order to determine if LA is a unique case that is skewing the data or if it is the most obvious example of a common trend in the data that could explain the decrease in solar households. The first test done was to determine if other highly urbanized counties had suffered from a loss in solar households and if any patterns could be found there. This was done using population density as a

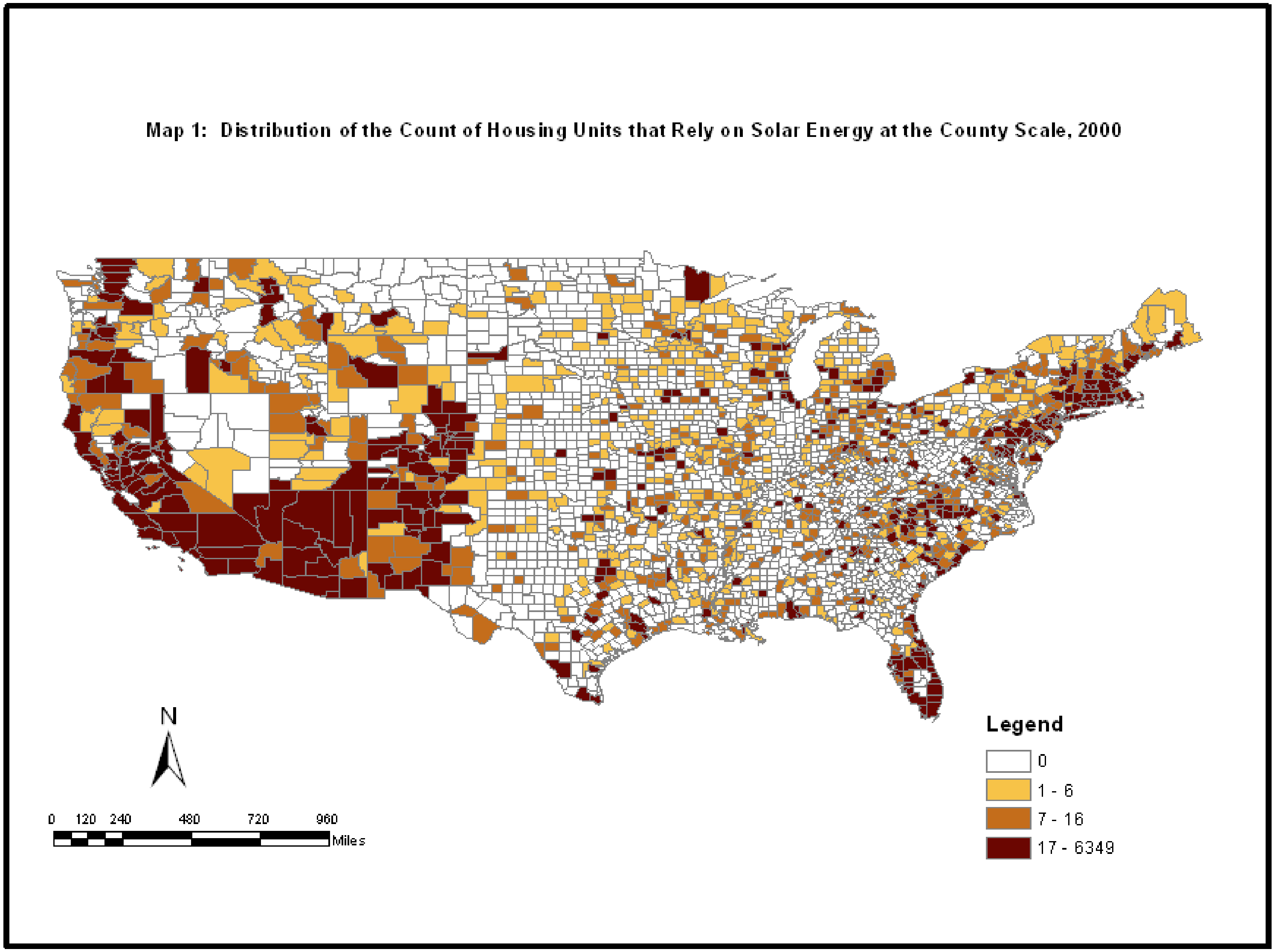

Figure 1:Distribution of the Count of Housing Units that Rely on Solar Energy at the County Scale, 2000

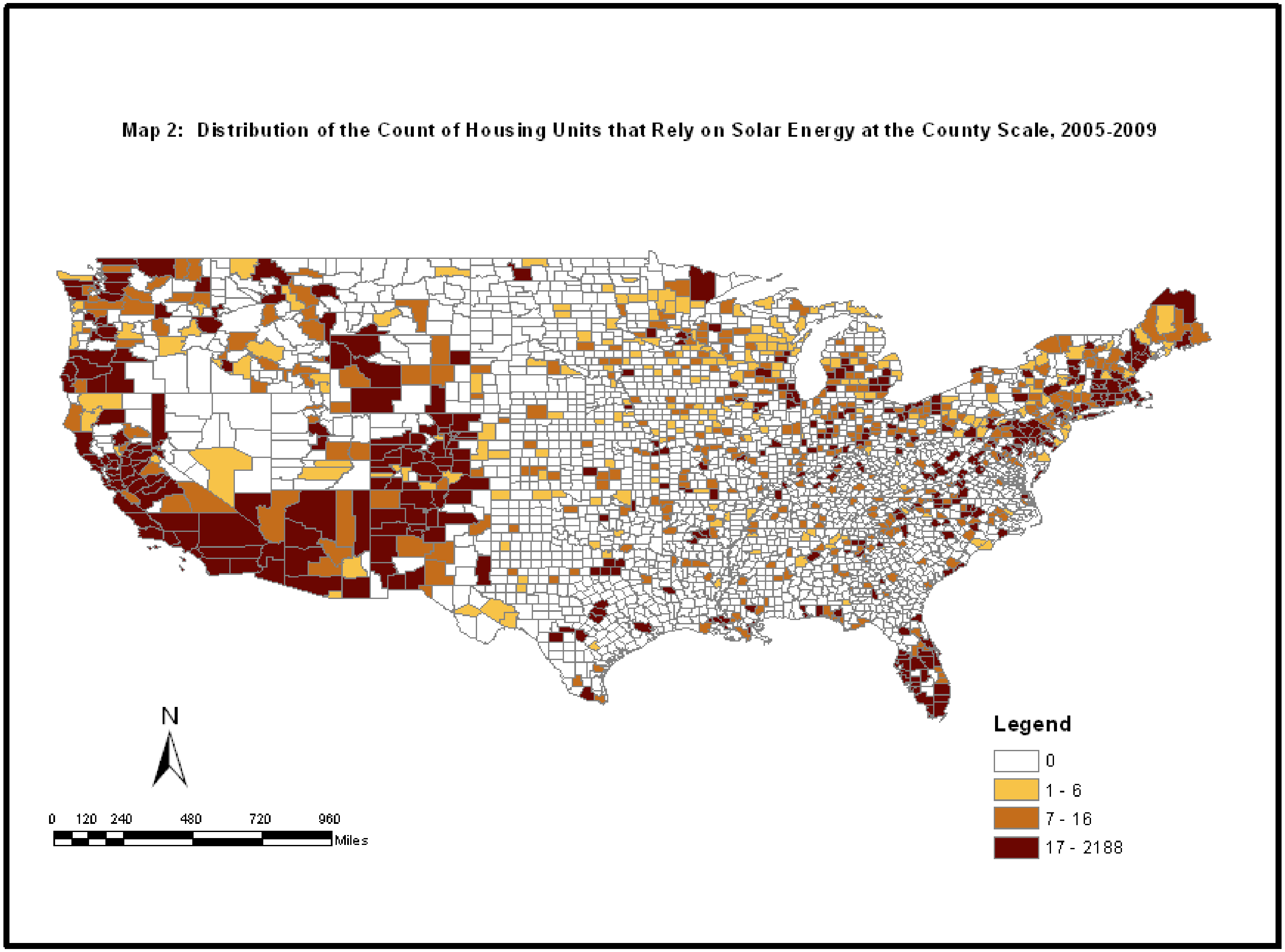

Figure 2: Distribution of the Count of Housing Units that Rely on Solar Energy at the County Scale, 2005-2009

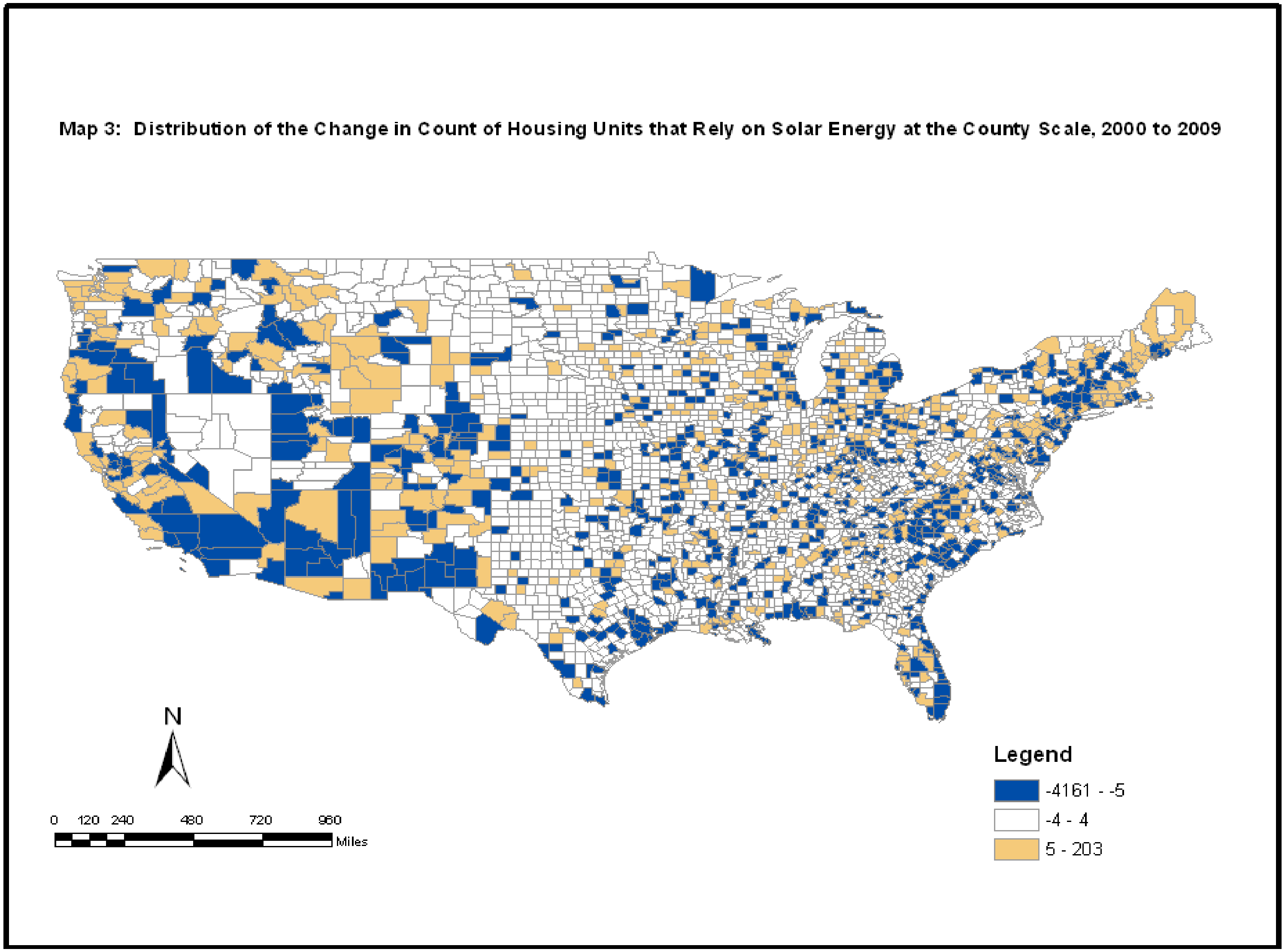

Figure 3:Distribution of the Change in Count of Housing Units that Rely on Solar Energy at the County Scale, 2000 to 2009

measure for the urbanization of a county and by creating a new variable called *solar diff* that measures the difference in the number of occupied housing units that rely on solar heating from 2000 to 2005-2009. By splitting the counties into quantiles (after removing LA county) based on their population density and then taking the summary statistics of *solar diff* of each quantile it is easy to see if the drop in solar households came from one specific quantile of counties or if it was spread relatively evenly between all of them. Table 5 below does just that using 16 quantiles.

Table 5. Difference in the number of occupied housing units that rely on solar heating from 2000 to 2005-2009 sorted by population density quantiles (excluding LA county)

Pop density Quantiles	Observations	Mean	Standard Deviation	Minimum	Maximum
1	198	-0.10606	4.518868	-29	37
2	196	1.091837	8.063322	-21	47
3	194	-0.7268	13.67736	-105	124
4	187	1.754011	10.34218	-30	76
5	189	0.587302	7.568063	-18	57
6	193	-0.64249	8.42687	-44	57
7	186	-0.94624	11.94661	-79	56
8	184	-0.34239	9.048572	-56	34
9	194	0.273196	8.717953	-22	55
10	193	-1.48705	10.98536	-58	48
11	194	-1.42268	28.16967	-336	94
12	196	-1.18878	20.94021	-243	63
13	198	-1.25758	13.98183	-87	52
14	193	1.507772	18.49863	-52	100
15	205	-4.1122	35.64974	-321	145
16	206	-20.6311	92.67251	-896	203

Table 5 clearly shows a large drop in the mean and increase in the standard deviation for the final quantile, which is made up of counties with population densities of at least 600, compared to any of the previous quantiles; although the 15th quantile shows the beginning of this trend, with a similar, if much smaller, decrease in mean and increase in the standard deviation. In order to even more clearly show the relationship between the decrease in the number of occupied housing units that rely on solar heating and densely populated counties Table 6 shows the total difference in solar households compared to the count of solar households in 2005-2009, again sorted by population density quantiles.

Table 6. Total Solar difference and Solar count by Population Density Quantile

Pop. density Quantiles	Total solar difference	Std. Err.	Solar count 2005-2009	Std. Err.
1	-21	63.58611	268	73.5293
2	214	112.8865	799	158.0357
3	-141	190.5035	704	211.8837
4	328	141.4272	1085	433.4517
5	111	104.0437	737	206.8756
6	-124	117.0698	512	167.4966
7	-176	162.93	749	189.7854
8	-63	122.7408	644	123.0704
9	53	121.427	663	111.5825
10	-287	152.6136	927	189.591
11	-276	392.3581	1575	522.9435
12	-233	293.1629	1470	367.6884
13	-249	196.7418	1746	348.3105
14	291	256.9911	2286	313.4877
15	-843	510.4266	5064	819.6598
16	-4250	1330.101	9696	1438.828

Table 6 shows that, while the last quantile has a higher solar household count than any other quantile, it still suffered a disproportionately larger loss in solar households than any other quantile. This reinforces the results from table 6, making it clear that LA was not a unique case, but that something is happening in densely populated counties that have a significant, negative impact on the number of occupied housing units that rely on solar heating in those counties.

Regression Results

The ZINB regression model was built following Zahran's (2008) example and so the independent variables were added in increments to the model. This is done to provide transparency to the model and demonstrate the robustness of the regression by showing that meaningful results can be obtained even when varying which variables are included. The baseline model is the same, using only the number of occupied housing units to predict the number of housing units that rely on solar heating. The pre/post time variable is included in all the other models. So the next model has both a time variable and environmental variables. For Model 3 both the economic and sociopolitical variables were added. In Model 4, financial incentive and regulatory variables are added. Thus model 4 is the complete model with all the variables. Table 7 contains the coefficients, standard errors, and statistical significance of the variables in each model as well as alpha, which is a measure for the over-dispersion in the data.

Table 7. Zero-inflated negative binomial regression models predicting housing units per count with solar heating

	Model 1		Model 2		Model 3		Model 4	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Negative binomial portion								
Constant	2.47**	0.031	-13.93**	2.45	-16.32**	2.32	-16.82**	2.32
Housing units	0.00526**	0.00024	0.0047**	0.0002	0.0027**	0.0002	0.0027**	0.0002
pre/post			0.2186**	0.0406	0.2061**	0.0418	0.4995**	0.0909
Solar radiation			0.4092**	0.0227	0.4353**	0.0228	0.4179**	0.0236
Mean temp min			0.0368**	0.0025	0.0390**	0.0026	0.0379**	0.0027
Mean temp max			0.4439**	0.0651	0.4592**	0.0617	0.4681**	0.0617
Mean temp max squared			-0.0035**	0.0004	-0.0035**	0.0004	-0.0035**	0.0004
Median home value					-0.0010*	0.0005	-0.0005	0.0005
Median household Income					0.0222**	0.0036	0.0192**	0.0036
Solar providers					0.1434*	0.0716	0.1367ψ	0.0715
Population density					2.59E-5**	8.78E-6	2.19E-5*	8.85E-6

Consumption Age					1.9728	1.4578	1.9964	1.4583
Net democrat					0.9351**	0.0858	0.9162**	0.0874
Environmental Groups					0.0397	0.0523	0.0409	0.0525
ICLEI participation					0.3306**	0.0748	0.3225**	0.0743
Total Financial incentive							-0.0039	0.0077
Total Solar regulation							0.0266**	0.0094
Financial difference estimator							-0.0021	0.0115
regulatory difference estimator							-0.0406**	0.0137
Logistic portion								
Constant	1.29**	0.05	-0.30	5.14	8.03	5.23	11.51*	5.12
Housing units	-0.0527**	0.0034	-0.0498**	0.0031	-0.0373**	0.0028	-0.0368**	0.0027
pre/post			0.8268**	0.0666	1.0350**	0.0721	1.2759**	0.1365
Solar radiation			-0.3132**	0.0451	-0.3269**	0.0490	-0.2961**	0.0491
Mean temp min			-0.0087ψ	0.0046	-0.0012	0.0055	-0.0026	0.0054
Mean temp max			-0.0429	0.1390	-0.2370ψ	0.1405	-0.3213*	0.1374
Mean temp max squared			0.0011	0.0010	0.0021*	0.0010	0.0026**	0.0009
Median home value					-0.0076**	0.0016	-0.0060**	0.0015
Median household Income					0.0066	0.0078	0.0074	0.0077
Solar providers					-0.2501	0.3551	-0.3028	0.3577
Population density					0.0004**	0.0001	0.0004**	0.0001
Consumption Age					1.6565	2.3990	1.4907	2.3920
Net democrat					-0.9275**	0.1500	-0.7491**	0.1547
Environmental Groups					-0.2724	0.2120	-0.2996	0.2116
ICLEI participation					-0.3833	0.2403	-0.4422 ψ	0.2424
Total Financial incentive							0.0050	0.0158
Total Solar regulation							-0.0410*	0.0160
Financial difference estimator							0.0232	0.0232
regulatory difference estimator							-0.0629**	0.0237
lnalpha	0.2975**	0.0436	-0.1542**	0.0408	-0.3509**	0.0401	-0.3663**	0.0396
alpha	1.3465	0.0587	0.8571	0.0350	0.7040	0.0283	0.6933	0.0275

Ψ$p<.10$ *$p<.05$ **$p<.01$

Housing units Median home value and Median household Income are measured in thousands

Table 8 converts these coefficients into the expected percentage change in the number of housing units with solar heating both for a one unit change and for a one standard deviation change in each variable in each model using the listcoef command developed for Stata by Long and Freese. Table 8 also reports the fit statistics for each of the models.

Table 8. Percentage of change in housing units with solar heating

	Model 1		Model 2		Model 3		Model 4	
	per unit change	per SD change	per unit change	per SD change	per unit change	per SD change	per unit change	per SD change
Negative binomial portion								
Housing units	0.5	75.4	0.5	64.0	0.3	32.9	0.3	32.7
pre/post			24.4	11.5	22.9	10.9	64.8	28.4
Solar radiation			50.6	38.3	54.5	41.1	51.9	39.2
Mean temp min			3.8	55.7	4.0	59.8	3.9	57.7
Mean temp max			55.9	975.4	58.3	1066.8	59.7	1123.6
Mean temp max squared			-0.4	-94.1	-0.3	-93.9	-0.4	-94.0
Median home value					-0.1	-6.3	-0.0	-3.0
Median household Income					2.2	22.5	1.9	19.2
Solar providers					15.4	2.8	14.6	2.6
Population density					0.0	4.5	0.0	3.8
Consumption Age					619.1	3.2	636.3	3.2
Net democrat					154.7	28.2	150.0	27.5
Environmental Groups					4.0	1.1	4.2	1.1
ICLEI participation					39.2	6.3	38.1	6.1
Total financial incentives							-0.4	-1.4
Total solar regulations							2.7	10.0
Financial difference estimator							-0.2	-0.7
regulatory difference estimator							-4.0	-14.6
Logistic portion								
Housing units	-5.1	-99.6	-4.9	-99.5	-3.7	-98.1	-3.6	-98.0
pre/post			128.6	51.2	181.5	67.8	258.2	89.3
Solar radiation			-26.9	-22.0	-27.9	-22.8	-25.6	-20.9
Mean temp min			-0.9	-9.9	-0.1	-1.4	-0.3	-3.1
Mean temp max			-4.2	-20.5	-21.1	-71.9	-27.5	-82.1
Mean temp max squared			0.1	135.6	0.2	435.1	0.3	705.1
Median home value					-0.8	-37.9	-0.6	-31.3
Median household Income					0.7	6.2	0.7	7.0
Solar providers					-22.1	-4.6	-26.1	-5.6
Population density					0.0	99.6	0.0	91.5
Consumption Age					424.1	2.7	344.0	2.4
Net democrat					-60.4	-21.8	-52.7	-18.0
Environmental Groups					-23.8	-7.3	-25.9	-8.0
ICLEI participation					-31.8	-6.8	-35.7	-7.8
Total financial incentives							0.5	1.8
Total solar regulations							-4.0	-13.6
Financial difference estimator							2.3	8.4
regulatory difference estimator							-6.1	-21.7

Total observations	6217	6213	6212	6212
Nonzero observations	2354	2351	2351	2351
Zero observations	3863	3862	3861	3861
Vuong z	22.09	27.39	28.32	28.91
Prob. > z	0.0	0.0	0.0	0.0
LR	2581.31	3659.54	4051.91	4108.07
Prob. > LR	0.0	0.0	0.0	0.0
McFadden's Adj R^2	0.092	0.13	0.145	0.144
Maximum likelihood R^2	0.34	0.445	0.479	0.484
Cragg and Uhler's R^2	0.344	0.45	0.484	0.489

For each of the models there are two separate portions; one that estimates the count of housing units with solar heating and one that estimates the probability that a county will contain zero housing units with solar heating. This means that a negative value indicates that a variable is decreasing the odds that there will be zero housing units with solar heating in that county. For each model the effect that each variable has on both portions will be analyzed.

Model 1

To start off in the baseline model the only independent variable is the number of *occupied housing units*. The result is that the expected count of housing units with solar heating increases by 75.4% with a one standard deviation increase in *housing units*. In the logistic portion a one standard deviation increase in *housing units* decreases the odds of having zero housing units with solar heating by 99.6%.

Model 2

When the time variable and the environmental variables are added in model 2 the result is that the expected count only increases by 64% for a one standard deviation increase in *housing units*. This indicates that the effect of *housing units* on the expected count was overstated in model 1 as the impact of the new variables were in Model 1 attributed to *housing units*. On the other hand a one standard deviation increase in *housing units* decreases the odds of zero by 99.5%, almost identical to the first model. This means that *housing units* was not absorbing the effects of the new variables had on the odds of a zero count.

Looking at the *pre/post* time variable, in the later time period there is a 24.4% increase in the expected count. Looking at the logistic portion however the later time period increases the odds of zero by 128.6%. *Solar radiation* increases the expected count by 38.3% for a one standard deviation increase and decreases the odds of zero by 22% for a one standard deviation increase. *Mean minimum temperature* increases the expected count by 55.7% with a one standard deviation increase and decreases the odds of zero by 9.9%. However *mean minimum temperature* is only statistically significant given a 10% margin of error in the logistic portion, while the previous variables were statistically significant given a 1% margin of error in the logistic portion. The expected count increases by 975.4% with a one standard deviation increase of the *mean maximum temperature* however it is statistically insignificant in the logistic portion. A one standard deviation increase in the *mean maximum temperature squared* decreases the expected count by 94.1% and is also statistically insignificant in the logistic portion. All variables were statistically significant given a 1% margin of error in the negative binomial portion.

Model 3

With the addition of the economic and sociopolitical variables the explanatory power of *housing units* dropped significantly. Now a one standard deviation increase in *housing units* only increases the expected count of housing units with solar heating by

33%. However its ability to predict whether a county will have zero housing units with solar heating has barely dropped at all with it decreasing the odds of such an event by 98.2%. *Pre/post* had little change, with the later time period increasing the expected count by 22.9% and is just as statistically significant as before. In the logistic portion the later time period increases the odds of zero by 181.5% so it has an even bigger impact than before. *Mean minimum temperature, mean maximum temperature, and mean maximum temperature squared* all remain roughly the same as in Model 2, in the negative binomial portion. In the logistic portion *mean minimum temperature* is no longer statistically significant while *mean maximum temperature* is statistically significant given a 10% error margin and *mean maximum temperature squared* is statistically significant given a 5% error margin. A one standard deviation increase in *mean maximum temperature* decreases the odds of zero by 71.9% while a one standard deviation increase in *mean maximum temperature squared* increases the odds of zero by 435.1%.

For the economic variables it turns out that a one unit increase ($1,000) in *median home value* decreases the expected count by .1% and decreases the odds of zero by .8%. It is statistically significant given a 5% error margin in the negative binomial portion and is statistically significant given a 1% error margin in the logistic portion. A one unit increase ($1,000) in *median household income* on the other hand increases the expected count by 2.2% and is statistically significant given a 1% margin of error but is not statistically significant in the logistic portion. The presence of *solar providers* increases the expected count by 15.4% and is statistically significant given a 5% margin of error and is not statistically significant in the logistic portion. A one standard deviation increase in *population density* increases the expected count by 4.5% and increases the odds of zero by 99.6%. It is statistically significant given a 1% margin of error in both the negative binomial and logistic portions. Finally *consumption age* is statistically insignificant in both portions of the model.

For the sociopolitical variables a one standard deviation increase in *net Democrats* increases the expected count by 28.2% and decreases the odds of zero by 21.8%. It is statistically significant given a 1% margin of error in both the negative binomial and logistic portions. *Environmental groups* are statistically insignificant for both portions of the model. *ICLEI* participation increases the expected count by 39.2% and is statistically significant given a 1% margin of error but is not statistically significant in the logistic portion.

Model 4

The addition of financial and regulatory variables has little impact on the effects of many of the variables from model 3. *Pre/post* now has a larger effect, with the later time period increasing the expected count by 64.8 and increasing the odds of a zero count by 258.2%. *Mean maximum temperature* and *mean maximum temperature squared* both increased in statistical significance in the logistic portion so now *mean maximum temperature* is statistically significant given a 5% margin of error and *mean maximum temperature squared* is statistically significant given a 1% margin of error. A one standard deviation increase in *mean maximum temperature* now decreases the odds of zero by 82.1% while a one standard deviation increase in *mean maximum temperature squared* now increases the odds of zero by 705.1%. *Median home value* is no longer statistically significant and *solar providers* and *population density* are both less statistically significant in the negative binomial portion while their effects remain roughly the same. *Solar providers* is now only statistically significant given a 10% margin of error and *population density* is now only statistically significant given a 5% margin of error. *Population density* also retains roughly the same effects and its statistical significance in the logistic portion. Finally *ICLEI* participation is now only statistically significant given a 10% margin of error in the logistic portion, with participation decreasing the odds of a zero count by 35.7%.

Neither t*otal financial incentives* nor f*inancial difference estimator* are statistically significant in either the negative binomial or the logistic portions. *Total solar regulation* is statistically significant given a 1% margin of error in the negative binomial section and is statistically significant given a 5% margin of error in the logistic portion. A one standard deviation increase in *total solar regulation* increases the expected count by 10.0% and decreases the odds of a zero count by 13.6%. The *regulatory difference estimator* is statistically significant given a 1% margin of error in both the negative binomial and logistic portions. For the *regulatory difference estimator* a one standard deviation increase in the *regulatory difference estimator* decreases the expected count by 14.6% and also decreases the odds of a zero count by 21.7%.

Fit of Models

As variables are added to each subsequent model the fit of the model improves as shown by the increase in the various pseudo R^2 measures; for example Cragg and Uhler's R^2 increases from .344 for model 1 to .45 for model 2 and .484 for model 3. Model 4 though has almost the exact same fit as model 3 with some of the R^2 measures staying the same and the rest only showing a marginal increase such as Cragg and Uhler's only increasing to .489. This, when combined with how the coefficients and standard deviations of the variables all remained fairly steady across all of the models, indicates the general robustness of Zahran et al.'s method.

Maps provide additional insight into the goodness-of-fit question. Figure 4 shows the distribution of the predicted count of housing units that rely on solar heating in 2005-2009. By comparing this map to Figure 2, which shows the distribution of the observed count, the model can be judged on how well Figure 4 matches Figure 2. While it looks like Figure 4 does a fairly good job of predicting the general areas, it appears to be over predicting the count in many areas, even after adjusting the first quantile to go from 0 to 2 instead of just 0 as in Figure 2. The problem is that the ZINB cannot properly report a zero count, except in the most extreme cases. This is because of how it calculates expected count: $E(\, y_i|x_i\, z_i) = \mu_i(1 - \psi_i)$ where $\mu_i = \exp\,(x_i\, \beta)$. So it takes the expected mean and multiplies it by 1 minus the probability of being in the always zero group. So unless the probability is 100%, which in this model none of the counties are, then the expected count will be some positive value, even if it is just .5. Therefore the model cannot properly report zero counts, but that is fairly easy to adjust for as Figure 4 shows. How well does the model do at predicting non-zero counts though?

Residuals are a good way to determine that, so Figure 5 plots the distribution of the standardized residuals for 2005-2009. The residuals are standardized using this formula: $\frac{observed - \mu_i}{\sqrt{Var\,(y_i \mid x_i\ ,z_i\,)}}$, where $Var\,(y_i \mid x_i\ ,z_i\,) = \mu_i(1 - \psi_i\,)\,[1 + \mu_i(\psi_i + \alpha\,)]$ (Zahran, 2008). Looking at Figure 5, it appears as though the model systematically overestimated the count, with a majority of the counties being over 1 standard deviation below the predicted count. However, if Figure 5 is compared with Figure 2 closely those same counties match up almost exactly with the zero count counties. A close examination of the data confirms this, with all but a dozen of the 2,136 counties that were over 1 standard deviation below the predicted count being counties with a zero count. This is further proof that aside from not being able to predict actual zero counts the model does an excellent job predicting the number of housing units that rely on solar heating in a county.

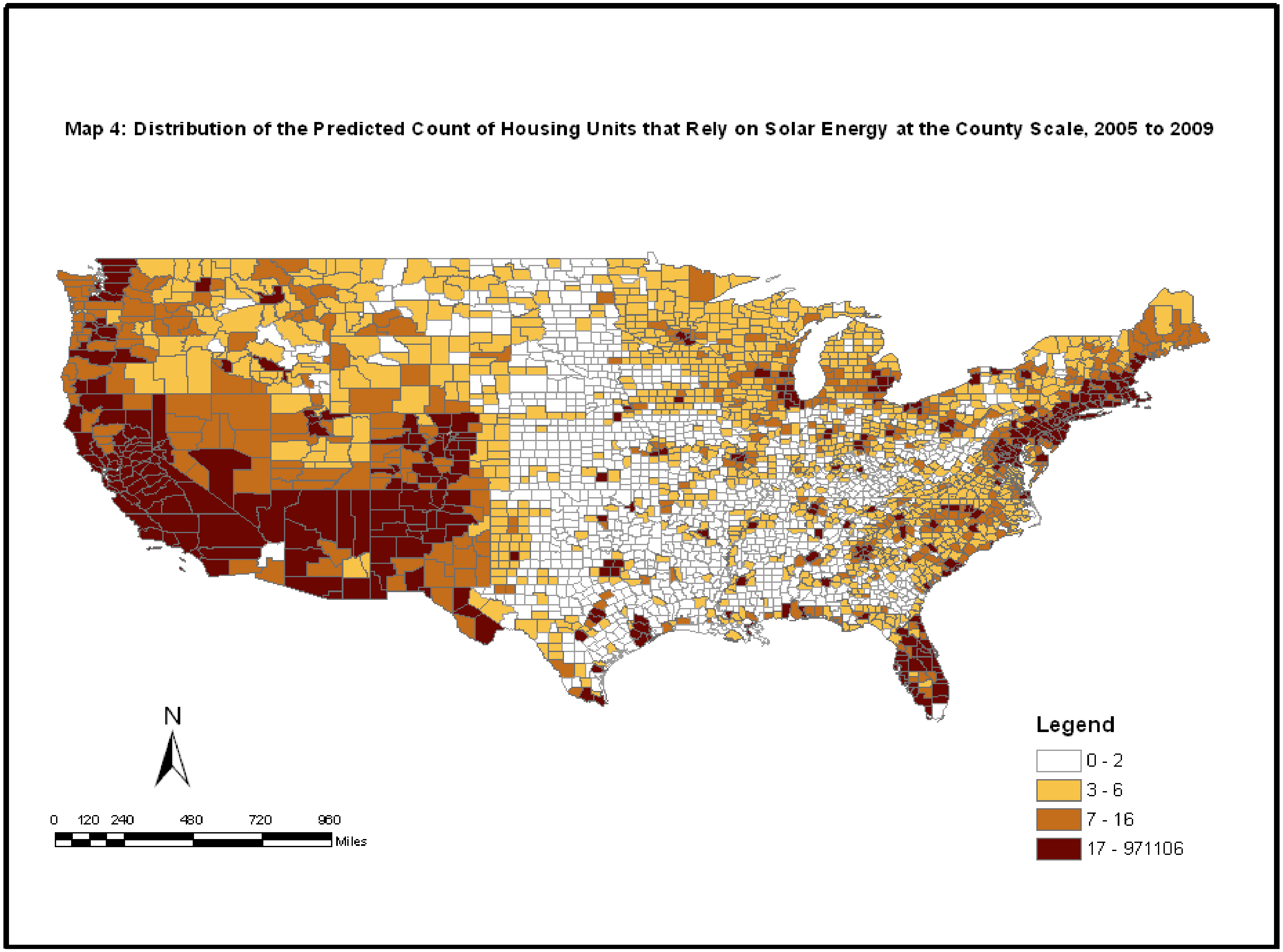

Figure 4: Distribution of the Predicted Count of Housing Units that Rely on Solar Energy at the County Scale, 2005-2009

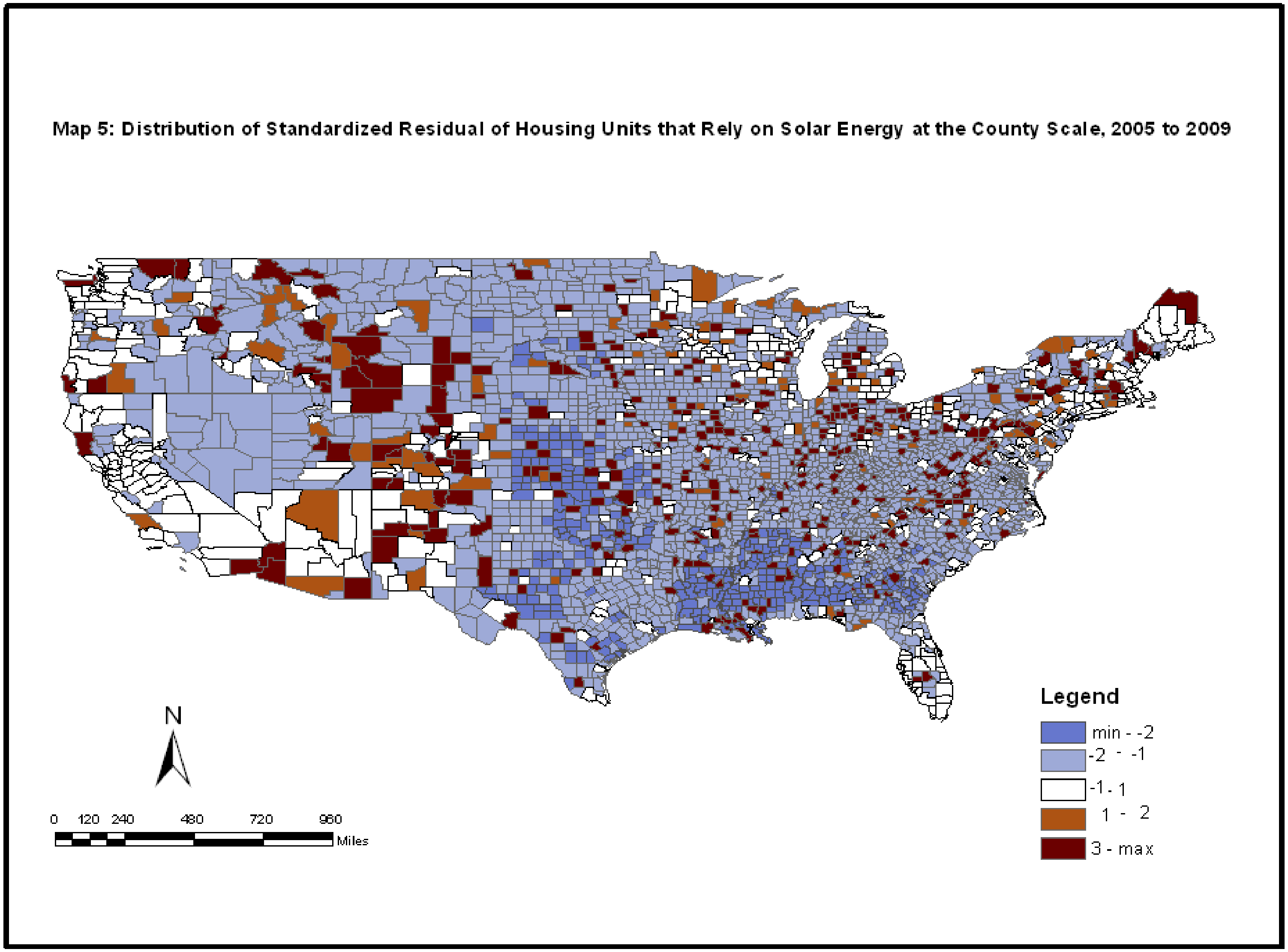

Figure 5: Distribution of Standardized Residual of Housing Units that Rely on Solar Energy at the County Scale, 2005-2009

Chapter 6: Discussion

Decrease in Solar Households

As was shown in tables 5 and 6, there is definitely a correlation between a densely populated county and a decrease in the number of occupied housing units that rely on solar heating in that county. This doesn't answer the question of what is causing that decrease however, instead merely providing a focus for future studies.

Original Variables

Now that the general validity and robustness of the ZINB model has been examined, the effects of the individual variables from model 4 can be examined and compared with their counterparts from Zahran et al.'s (2008) model. Zahran et al.'s (2008) findings have already been discussed in the literature review chapter and so will not be restated here.

A number of variables showed no appreciable change between the models. *Housing units, solar radiation, population density/urbanization, net Democrat, environmental groups, and ICLEI participation* all have the same sign and statistical significance. This means that Zahran et al.'s (2008) interpretations of their effects still is valid.

In terms of changes, the temperature variables both became more likely to be statistically significant, with both becoming statistically significant given a 1% margin of error in the negative binomial section, while keeping the same signs. More striking is the huge increase in the effect these variables have with *mean maximum temperature* now increasing the expected count by 1033.7% with a one standard deviation increase, while the original variable only increased it by 51.8%. The same happened for *mean maximum temperature square*, if not quite to the same degree, going from a 24% decrease to a 93.7% decrease. This is likely a result of switching to new temperature measurements that are more representative of the actual climate of an area. Zahran et al.'s (2008) interpretations of climate effects is thus still valid as these new climate variables simply provide a different, slightly better representation of the climate and that is behind the changes in the results and not a change in the relationship between climate and the adoption of solar heating systems.

Median home value went through a major change, with it no longer being statistically significant in the negative binomial portion. The loss of statistical significance is to be expected with the inclusion of *median household income* as a variable as home value was serving as something of a proxy for that measurement in Zahran et al.'s (2008) model. The variable kept the same sign, statistical significance, and general impact as in Zahran et al.'s (2008) model for the logistic portion.

The presence of *solar providers* was statistically significant in the new model, at least in regards to the expected count when it was statistically insignificant in Zahran et al.'s model. This is interesting for several reasons. First off is the fact that its percentage change for the expected count is very close to what it was for the original model. That similarity is likely due to the fact that this variable was not updated so it is using the 2000 values for both periods. The most likely explanation is that the solar providers that were a part of SEIA in 2000 are having a larger impact on consumers' decisions than they did before. This could be due to increased marketing or doing a better job of providing information and guidance than they did before or simply by providing better installation rates.

The final variable to have changed is *consumption age*. For some reason the fraction of the population in the ideal consumption age is no longer statistically significant. While there was an overall decrease in the fraction of the population that fell in that age due to the aging of the baby boomers, that should not affect significance, especially with the corresponding decrease in houses with solar heating. The best explanation is that the new group of citizens between the ages of 40 to 49 may not share the spending characteristics of the previous group. As a rough test of this theory a new variable was created that would capture as best it could the group of people who were 40 to 49 years old in 2000 in order to measure the impact of a cohort instead of an age group. To do this the data from the consumption variable for 2000 was kept and the percentage of the population between ages 50 to 59 for 2005-2009 from the ACS survey was added. While not a perfect fit this new variable now provides a rough estimate of how much of a presence that cohort has in a county. When the age variable in model 4 was replaced with this new cohort variable the new model suffered a very slight decrease in fit in some of the pseudo R^2 measures, but the new cohort variable was statistically significant in both the count and logistic portions with the expected signs. This matches with the original theory, but upon further reflection it might not be spending characteristics that are the major difference between the two groups. Rather it could be that the earlier cohort is more environmentally conscious than the later cohort.

New Variables

As one of the new variables *mean minimum temperature* helps paint a fuller picture of the impact of climate on solar energy adoption. Minimum temperature is statistically significant and has a positive impact on expected count. This supports the idea that solar heating systems fare poorly in especially cold climates. The *median household income* of a county is also statistically significant and has a positive impact on the expected count. This makes sense as a solar heating system is an expensive investment so a household's income would be an important deciding factor.

Finally there are the financial and regulatory variables whose analysis is one of the main goals of this thesis. None of the financial incentive variables were statistically significant in either portion of the model. These results disagree with established literature, contradicting the results from many studies such as Hoffman and Kiefer (2001), Hasset (1993), and Hayne (2002) to name a few. The most likely reason for this disparity is that the data are too coarse. The first problem is that the data only goes to the state level and not the county level. This is due to the fact that there were too few observations of local level incentives or regulations to model. The biggest failing of the data though is that it makes no attempt to measure the size, value, or any other characteristics the might determine how successful an incentive or regulation might be.

On the other hand, both of the solar regulation variables are statistically significant in both portions of the model. *Total solar regulations* being statistically significant means that there are unobserved characteristics that play a role in determining both how many solar regulations a county will enact and the number of housing units that rely on solar heating. The *regulatory difference estimator* being statistically significant on the other hand means that solar regulations have an impact on the number of housing units that rely on solar heating and the odds of there being no housing units that rely on solar heating. The logistic portion is just as expected, with additional solar regulations decreasing the odds of there being zero housing units that rely on solar heating. On the other hand the negative binomial portion is unexpected as it has additional solar regulations decreasing the expected count of housing units that rely on solar heating. As with the financial incentives this could be the result of the data being too coarse, thus skewing the results. This coarseness of the data means that even though the regulatory variables are reported as being statistically significant, it would be premature to attach any meaning to those results as the data is not accurately capturing the effects of solar regulations.

Chapter 7: Conclusions

This thesis set out to achieve two major objectives, with a third objective added at the end. The first was the update and analysis of the ZINB model used by Zahran et al. (2008) first in regards to its validity and robustness and then in either confirming or noting any changes in its analysis of the effect of climate, economic, and sociopolitical factors on the count of households in a county using solar heating. The second objective was to use the model to provide an empirical measure of the effect financial incentives and solar regulations have on the count of solar using households. The final objective was to explore and explain the unexpected decrease in residential solar usage.

The first objective was met, with the model, despite a few quirks, proving to accurately predict the number of occupied housing units that rely on solar heating in a county. While the effects of some variables had changed between the new model and Zahran et al.'s (2008) model, for the most part the original conclusions drawn by Zahran et al. still hold. From a policy making standpoint the only noteworthy changes were the presence of solar providers becoming statistically significant for determining the expected count of housing units that rely on solar heating and ICLEI participation becoming statistically significant for decreasing the odds of there being zero housing units that rely on solar heating. This is because policies can be made to encourage solar providers to set up business within a county and of course the local government can decide to join the ICLEI. (Consumption Age- add it)

There was less success in meeting the second objective, however. Measures of financial incentives turned out to be statistically insignificant while solar regulations had contradictory results. Therefore, few recommendations can be made based on those results. There are some policy implications to consider, despite all the problems. What this does show is that just introducing a financial incentive or solar regulation is not enough to guarantee an increase in housing units that rely on solar heating. This means that the incentive or regulation should be tailored to best fit the specific area it is to be implemented. This ties back to the first goal as the climate, economic, and sociopolitical characteristics of a county can be used to determine what type of incentive or regulation should be used. For example counties with high economic characteristics, but low sociopolitical characteristics should focus on incentives and regulations that build environmental consciousness instead of focusing on making solar systems more affordable.

The final objective was mostly successful. There was a clear correlation between the drop in residential solar usage and densely populated counties that will provide a sound starting point for any future studies on the issue. While there is too here to offer suggestions on how to address this decrease it is now clear where any efforts to do so need to take place.

There are a number of things that this thesis did not cover or covered inadequately that any future research should cover. The biggest area to expand on is to obtain more meaningful measurements for financial incentives and regulations so that their impact can be better studied. Along those same lines updating the solar providers and urbanization variables should increase the explanatory powers of the model. The model could also be expanded with more climate, economic, and sociopolitical variables to improve the explanatory power of the model and provide policy-makers a more complete picture. The most important variables to add would be local cost variables. A more in-depth study of the correlation between the decrease in housing units that rely on solar heating and densely populated counties should also be done. Given the apparent clustering of counties regarding solar adoption levels, a spatial econometric model could be applied to explore whether there is spatial autocorrelation, that is, a “neighbor effect” of the sort found by Gillingham (2010) and Rothfield (2010). Finally, if the data can be found, a similar model should be built that measures the number of solar energy systems instead of solar heating systems.

Works Cited

Bollinger, B., & Gillingham, K. (2010, August). Environmental Preferences and Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Marketing Science* .

Erge, T., Hoffman, V., & Kiefe, K. (2001). The German Experience with Grid-connected PV-systems. *Solar Energy , 70*, 479-487.

Faiers, A. (2009, May). Understanding the Adoption of Solar Power Technologies in the UK Domestic Sector. Cranfield University.

Gordon, P., & Parker, J. (2002). Consumption over the life cycle. *Econometrica , 70* (1), 47-89.

Gouchoe, G., Everette, V., & Haynes, R. (2002). *Case Studies on the Effectiveness of State Financial Incentives for Renewable Energy.* Golden: Nation Renewable Energy Laboratory.

Haas, R., Ornetzed, M., Hametner, K., Wroblewski, A., & Hubner, M. (1999). Socio-Economic Aspects of the Austrian 200 kWp-Photovoltaic-Rooftop Programme. *Solar Energy , 66*, 183-191.

Hasset, K. A., & Metcalf, G. E. (1995). Energy tax credits and residential conservation investment: Evidence from panel data. *Journal of Public Economics , 57*, 201-217.

Holbert, K. E. (2007). An Analysis of Utility Incentives for Residential Photovoltaic Installations in Phoenix, Arizona. *North American Power Symposium* (pp. 189-196). La Cruces: IEEE.

Long, J. E. (1993). An econometric analysis of residential expenditures on energy conservation and renewable energy sources . *Energy Economics* , 232-238.

Long, J. S., & Freese, J. (2005). *Regression Models for Categorical Dependent Variables Using Stata* (2nd Edition ed.). College Station, TX: Stata Press.

Painuly, J. (2001). Barriers to renewable energy penetration; a framework for analysis. *Renewable Energy , 24*, 73-89.

Payne, A., Duke, R., & Williams, R. H. (2001). Accelerating residential PV expansion: supply analysis for competitive electricity markets. *Energy Policy , 29*, 787-800.

Perez, R., Burtis, L., Hoff, T., Swanson, S., & Herig, C. (2004). Quantifying residential PV economics in the US—payback vs cash flow determination of fair energy value. *Solar Energy , 77*, 363-366.

Rothfield, E. (2010). Solar Photovoltaic Installation in California: Understanding the Likelihood of Adoption Given Incentives, Electricity Pricing and Consumer Characteristics. Durham: Duke Univeristy.

Zahran, S., Brody, S. D., Vedlitz, A., Lacy, M. G., & Schelly, C. L. (2008). Greening Local Energy: Explaining the Geographic Distribution of Household Solar Energy Use in the United States. *Journal of the American Planning Association , 74*, 419-434.

Zahran, S., Brody, S. D., Grover, H., & Vedlitz, A. (2006). Climate change vulnerability and policy support. Society & Natural Resources, 19 (9), 771–789.

Appendix 1

Scoring system for state support of residential solar renewable technologies.

Based on analysis of the DSIRE database (Alaska and Hawaii not included).

1. Financial incentives.

DSIRE provides an index of 10 financial incentive systems for the support of renewable energy. Five of the categories are designed for non-residential systems (corporate tax, industry support, bonds, production incentives, grants). Because the focus of the analysis is on residential use, these five categories were dropped, leaving five. The rebate and loan categories were both divided by either state or utility sponsorship, resulting in a total of seven scoring categories. See Table 1 for a description.

The DSIRE database may only be searched using one conditional specification at a time within either renewable or efficiency programs. It was therefore necessary to reexamine each of the 468 entries in the DSIRE summary table to identify solar-specific residential incentives, and then to differentiate solar water heating and solar space heating as subcategories of all solar incentive programs. Solar space heating programs were also scored separately if they were in place in 2000. Thus, four variables were scored for each incentive: total solar (all forms), programs including solar space heat, programs including solar space heat in 2000, and programs including solar water heating (scores don't total due to overlap). Binary scoring was used to indicate whether or not the incentive was provided in any form at the state level. Reducing the measure to binary was especially desirable due to the manner in which many states are served by multiple utility companies, each of which might offer its own rebate or loan program. Counting each of these instances will inflate the score for larger states. The four state-level financial incentives indices are the sum of the binary scores over the seven categories.

While the majority of the programs are at the state level, within the DSIRE database there are 47 instances in which local programs are identified (after being reduced to residential solar only). Each of these was examined to determine if a specific county or counties could be assigned. It was also possible to assign some of the utility-sponsored incentive programs to a local level. When examination of a utility-sponsored program clearly indicated an exclusive county or municipal attachment then this entry was scored only at the local level. This analysis yielded 62 counties with incentives above those scored at the state-level. These indices were summed across six categories (dropping state income tax), with differentiation between all solar, including solar space heat, and including solar hot water. No local programs were identified that were in place in 2000.

2. Regulatory environment.

The DSIRE database describes 12 regulatory mechanisms that support renewable energy, as described in Table 2. These can not be easily separated into mechanisms that apply only to the residential sector. It is possible to identify subsets of the regulatory categories that are most germane to solar space heating and solar water heating (e.g., net metering is not, solar access is). Two separate indices were therefore calculated, again using a binary scoring scheme to indicate whether or not the regulatory mechanism exists in some form at the state level. An index of five items was summed for regulations that are germane to solar space heat and hot water, and an index of 12 items was summed over the full set of regulations as an indicator of the state's general regulatory maturity for support of renewable energy. Again, separate measures were made for all regulatory mechanisms in place in 2000 and the subset representing heat- or water-relevant programs that were in place in 2000.

The DSIRE summary table of regulatory mechanisms also indicates the presence of local programs. A total of 85 such instances were identified in the summary table, representing counties in 23 states and eight of the regulatory categories. Each was examined to determine if a specific county could be assigned. A total of 60 local regulations were identified as relevant to renewable sources, spread across 21 states and seven of the regulatory mechanisms (four germane to solar space heat and water heat). As above, two indices were summed for the total number of regulations and those germane to solar space heat and water. No local programs were identified that were in place in 2000 (partially due to this information being unavailable in many of the local regulatory descriptions in DSIRE).

Table 9. National Prevalence and Description of State-Level Financial Incentives Scoring Categories

from: (http://www.dsireusa.org/glossary/glossary.cfm?&CurrentPageID=8&EE=1&RE=1)

Mechanism	Prevalence	Description
Loan Programs	14 State 9 Utility	Loan programs provide financing for the purchase of renewable energy or energy efficiency systems or equipment. Low-interest or zero-interest loans for energy efficiency projects are a common demand-side management strategy for electric utilities. State governments also offer low-interest loans for a broad range of renewable energy and energy efficiency measures.
Rebate	19 State	States, local governments and utilities offer rebates to promote the

Programs	15 Utility	installation of renewable energy systems and energy efficiency measures. The majority of rebate programs that support renewable energy are administered by states, municipal utilities and electric cooperatives; these programs commonly provide funding for solar water heating and/or photovoltaic (PV) systems.
Personal Tax Incentives	17	Personal tax incentives include personal income tax credits and deductions. Many states offer these incentives to reduce the expense of purchasing and installing renewable energy or energy efficiency systems and equipment.
Property Tax Incentives	28	Property tax incentives include exemptions, exclusions and credits. The majority of property tax incentives provide that the added value of a renewable energy system is excluded from the valuation of the property for taxation purposes.
Sales Tax Incentives	13	Sales tax incentives typically provide an exemption from the state sales tax (or sales and use tax) for the purchase of a renewable energy system, an energy-efficient appliance, or other energy efficiency measures.

Table 10. National Prevalence and Description of State-level Regulation Scoring Categories
from: (http://www.dsireusa.org/glossary/glossary.cfm?&CurrentPageID=8&EE=1&RE=1)

Mechanism	Prevalence	Description
		Most Germane to Solar Space and Hot Water Heating
Public Benefit Funds	18	State-level programs typically developed during electric utility restructuring by some states in the late 1990s to ensure continued support for renewable energy resources, energy efficiency initiatives and low-income energy programs.
Contractor Licensing	9	Several states have adopted contractor licensing requirements for solar water heating, active and passive solar space heating, solar industrial process heat, solar-thermal electricity, and photovoltaics. These are designed to ensure that contractors have the necessary experience and knowledge to install systems properly.
Equipment Certification Requirements	3	Policies requiring renewable energy equipment to meet certain standards serve to protect consumers from buying inferior equipment. These requirements not only benefit consumers; they also protect the renewable energy industry by making it more difficult for substandard systems to reach the market.
Solar Access Laws	34	Solar access laws are designed to protect a consumer's right to install and operate a solar energy system at a home or business. Some solar access laws also ensure a system owner's access to sunlight. In some states, access rights prohibit homeowners associations, neighborhood covenants or local ordinances from restricting a homeowner's right to use solar energy.
Permitting	24	Permitting standards can facilitate the installation of wind and solar energy systems by specifying the conditions and fees involved in project

Standards		development. Some local governments have adopted simplified or expedited permitting standards for wind and/or solar. Also includes building energy standards (private and public).
		All Others
Generation Disclosure Rules	24	Some states require electric utilities to provide their customers with specific information about the electricity that the utility supplies. This information, which must be shared with customers periodically, usually includes the utility's fuel mix percentages and emissions statistics.
Renewables Portfolio Standards	32	Renewable portfolio standards require utilities to use renewable energy or renewable energy credits to account for a certain percentage of their retail electricity sales – or a certain amount of generating capacity – within a specified timeframe.
Net Metering	44	For electric customers who generate their own electricity, net metering allows for the flow of electricity both to and from the customer. During times when a customer's generation exceeds the customer's use, electricity from the customer flows back to the grid, offsetting electricity consumed by the customer at a different time.
Interconnection Standards	38	Interconnection standards govern the technical and procedural process by which an electric customer connects an electric-generating system to the grid. Interconnection standards specify the technical, contractual, metering, and rate rules that system owners and utilities must abide by.
Line Extension Analysis	3	When a prospective electric customer requests service for a home or facility that is not currently serviced by the grid, the customer usually must pay a distance-based fee for the cost of extending power lines. In many cases, it is cheaper to use an on-site renewable energy system. Certain states require utilities to provide information regarding renewable energy options when customers request a line extension.
Green Power Purchasing	10	Many state and local governments have committed to buying green power to account for a certain percentage of their electricity consumption.
Mandatory Utility Green Power Option	8	Several states require certain electric utilities to offer customers the option of buying electricity generated from renewable resources. Typically, utilities offer green power generated using renewable resources that the utilities own, or they buy renewable energy credits from a provider certified by a state public utilities commission