RESEARCH REPORT SERIES
(Statistics #2009-03)

Population Estimation Using Tract Level Geography and Spatial Information

Patrick M. Joyce[1]
Donald Malec

[1] University of Connecticut

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: February 27, 2009
Revised and reissued: September 30, 2009

Population Estimation Using Tract Level Geography and Spatial Information*

Patrick M. Joyce
University of Connecticut

Donald Malec
U.S. Census Bureau

Abstract

Surveys such as the American Community Survey provide samples that cover the country's geography more evenly than highly clustered personal interview surveys. As a result, small areas such as census tracts will usually contain sample both within a tract and also within nearby tracts. The potential for using nearby data to increase the efficiency of tract level estimates, with spatial models, is investigated for a group of census tracts in the Delmarva Peninsula. Using Census 2000 counts of housing unit vacancy and occupancy rates and corresponding 1990 data as covariates, small area estimates of vacancy rate and person per housing unit are made incorporating conditional auto-regressive (CAR) spatial models.

The following issues related to making model-based estimates are addressed: fitting a parametric small area model to the housing unit data, comparison of CAR spatial models with a traditional hierarchal model without spatial components, the practical reduction in variance achieved by using a CAR spatial model and, lastly, the sampling properties of the small area estimates drawn from samples of 2000 census counts using both the CAR spatial model and the hierarchical model.

1 Introduction

The small areas of small area estimation are usually prespecified in advance and are typically administrative units, such as counties or states. The definition of the small area boundaries, however, may or may not correspond to the homogeneity of the population within the small area or the heterogeneity between neighboring small areas. There may, in fact, be larger geographic regions that contain small areas that share similarities, including similar outcomes to survey questions. In situations like this, data collected in a small area may provide some information about its neighboring small areas. The preferential use of nearby similar data, in small area estimation, can be achieved by including the possibility of spatial similarity in a statistical model. Whether looking at maps of small area outcomes or just using the corresponding estimates, spatial models can increase the precision when characteristics of nearby areas are related.

The use of spatial models for small area estimation can be found in Ghosh et al. (1998) and references therein. For a comprehensive account of small area estimation, see Rao (2003).

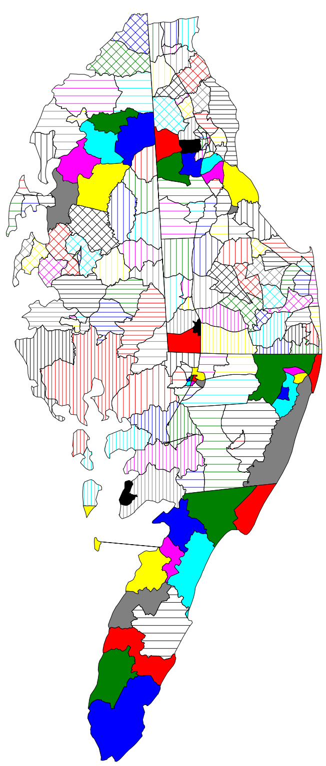
The use of spatial models at the Census Bureau is being evaluated at this time because data amenable to spatial modeling are now available from the American Community Survey (ACS). In order to economically spread the ACS sample over all census tracts in the United States, the U.S. mail is used as a first contact. Many other national surveys are collected via personal interviews that, to save costs, are collected within a sample of large primary sampling units (PSUs) typically made up of a county or a group of contiguous counties. PSUs are designed to reduce the total interviewer travel time. The use of any spatial model based on these small samples of PSUs would require extrapolation across large areas where the primary sampling units were not in sample. Telephone survey responses can be more geographically disperse and still be cost

*The authors thank Lynn Weidman (SRD) for many helpful discussions and comments throughout the project and Julie Tsay (SRD) for providing assitance with managing the data. In addition, the authors would like to thank Donald Luery (DID), April Avnayim (GEO) and Patrick Cantwell (DSSD) for providing comments on drafts of the report. Lastly, the authors thank Stephanie Sheffield (SRD) for editorial assistance.

effective. However, the use of different area codes or exchanges may only provide a rough idea of geographic proximity. Mail surveys, however, are linked to an address and, apart from Post Office boxes, can be linked to a precise geographic location.

This preliminary evaluation of the possible uses of spatial models for small area estimation for the ACS uses neither ACS data nor, for that matter, much data from the entire United States. Instead, Census 2000 data related to housing unit size and vacancy status will be used for making tract-level estimates for the Delmarva Peninsula. The reason Census data are used instead of ACS data is that the Census data provides a complete count, affording more data for model fitting and providing a large data source from which to sample in order to evaluate the proposed small area estimates. The census items used (housing characteristics) are also obtained in the ACS but are based on a different residence rule. The reason that only data from the Delmarva Peninsula are used is to avoid the need for extensive data processing. The Delmarva Peninsula, comprising Delaware and the eastern seashore parts of Maryland and Virginia, is about 5.45 thousand square miles in size, containing 165 census tracts (see fig. 1). The tracts are contiguous. Residences on the peninsula are occupied by both permanent and seasonal residents and should provide a variety of housing characteristics with which to model. Census tracts are picked as the small area because annual tract-level design-based estimates are not planned to be released due to their poor precision. Instead, ACS tract-level estimates are only planned for aggregate, multi-year design-based estimates. Previous work on model-based small area estimation for the ACS also focused on tract-level estimation (see Chand & Alexander (1995), Chand & Alexander (1996), Chand & Alexander (1999), Chand and Malec (2001) and Malec (2005)).

Figure 1: Census Tracts of the Delmarva Peninsula (Census 2000 definitions)



2 Overview of the Spatial Model Used and Its Non-spatial Counterpart

Our model can be broken into two components. The first component describes the distribution of outcomes at the tract level using a model with all tract-specific parameters. The second component describes how the tract-level parameters relate to one another. This second level model induces "borrowing strength," the signature of small area estimation.

2.1 Within Tract Level Model Component

For each housing unit, h, in tract, t, define X_{th} to be the housing unit size, $X_{th} \in \{0, 1, 2, \ldots\}$ (see section 3.1 for a detailed description of the data).

It is assumed that the X_{th} are distributed i.i.d. within tract t and that:

$$P(X_{th} = x | p_{tx}) = p_{tx}, \tag{1}$$

where, $x \in \{0, 1, 2, \ldots\}$ and $\sum_{x=0}^{\infty} p_{tx} = 1$.

Further, the following parameter transformations and parameter reduction are used:

$$\begin{aligned}
p_{t0} &= q_{t0} \\
p_{t1} &= (1 - q_{t0})q_{t1} \\
p_{t2} &= (1 - q_{t0})(1 - q_{t1})q_{t2} \\
p_{t3} &= (1 - q_{t0})(1 - q_{t1})(1 - q_{t2})q_{t3} \\
p_{t4} &= (1 - q_{t0})(1 - q_{t1})(1 - q_{t2})(1 - q_{t3})q_{t4} \\
p_{t5} &= (1 - q_{t0})(1 - q_{t1})(1 - q_{t2})(1 - q_{t3})(1 - q_{t4})q_{t4} \\
&\vdots \\
p_{t5+\ell} &= (1 - q_{t0})(1 - q_{t1})(1 - q_{t2})(1 - q_{t3})(1 - q_{t4})^{\ell-1}q_{t4} \\
&\vdots
\end{aligned} \tag{2}$$

The above parameter transformation is based on transforming the joint distribution of X_{th} in terms of a product of conditional distributions as follows:

$$P(X_{th} = x | X_{th} \geq x, q_{tx}) = \begin{cases} q_{tx} & 0 \leq \text{x} \leq 3 \\ q_{t4} & 3 < \text{x} \end{cases} \tag{3}$$

This conditional approach was used to transform a multinomial model of tract-level household composition in Chand and Malec (2001) and Malec (2005) because, based on this parameterization, the likelihood will factor into a product of single parameter terms. This parameterization is an extension of that idea to include an open-ended upper bound with the additional, data analytic observation, detailed in section 3.2, that constant conditional probabilities for all but a few values of x are justified.

The reason that a model was fit at the tract level in the first place was to avoid the underlying problems associated with the use of standard normal theory small area estimation. Specifically, to avoid the problems connected with assuming that tract-level statistics can be characterized as having a normal distribution and that the estimated tract-level variances are assumed to be error free and, hence, considered as known constants. As shown in Malec (2005), this assumption is untenable due to small sample size when applied to an example of ACS tract-level modeling. There, it was shown that the normality assumption could be replaced with a multinomial model which accounted for the sample design and, most importantly, accounted for the precision of all parameters in the model, including variance. For a recent account of the problems inherent in using the normality/precise-variance-estimates assumption for small areas, see Bell (2008).

For tract t, the likelihood function is:

$$\prod_{h \in S} P(X_{th} = x) \propto \quad \prod_{x=0}^{3} \quad q_{tx}^{f_{tx}} (1 - q_{tx})^{R_{tx}}$$
$$\times \quad q_{t4}^{\sum_{x=4}^{M_t} f_{tx}} (1 - q_{t4})^{\sum_{x=4}^{M_t} (x-4) f_{tx}}, \tag{4}$$

where, in tract t, M_t is the maximum housing unit size recorded, f_{tx} is the number of housing units recorded as size x and $R_{tx} = \sum_{\ell=x+1}^{M_t} f_{t\ell}$ is the number of housing units with a size recorded as larger than x.

2.2 Between-Tract-Level Model Component

The model specifying how the tract-level parameters, q_{t0}, q_{t1}, q_{t2}, q_{t3} and q_{t4}, are related across tracts is specified on the logistic scale, since all five parameters are between zero and one.

Define

$$logit(q_{ti}) = \alpha_i + x_{ti}\beta_i + e_{ti}, \quad i \in \{0, 1, 2, 3, 4\} \tag{5}$$

with possibly spatially correlated errors:

$$e_{ti}|e_{(-ti)} = \gamma_i \sum_{k \in B_t} e_{ki} + \epsilon_{ti}, \quad i \in \{0, 1, 2, 3, 4\} \tag{6}$$

where B_t denotes the set of tracts that are nearest neighbors to tract, t, and the notation: $y_{(-i)}$ denotes the set of all y_j such that $j \neq i$.

In addition,

$$\epsilon_{ti} \sim N(0, \sigma_i^2), \quad i \in \{0, 1, 2, 3, 4\}\ . \tag{7}$$

Note that complete independence is assumed between the i.

Issues in constructing covariates, x_{ti} and x_t are provided in section 3.3. A more detailed description of the spatial component of the model and alternative spatial models is provided in section 3.4.2.

This spatial model is compared with the "non-spatial" model implicitly defined from (6) by forcing $\gamma_i = 0$. In this case, a standard hierarchical model remains and equations (6) and (7) can be equivalently re-specified by the single equation:

$$e_{ti} \sim N(0, \sigma_i^2), \quad i \in \{0, 1, 2, 3, 4\}, \tag{8}$$

3 Choice of Model

The following details how the final model for evaluating the use of spatial models was selected. As will be seen, data analysis, covariate definition, and available spatial modeling software are all factors.

3.1 Data

In dealing with residential data our fundamental sampling unit is the housing unit. The response of interest from the housing unit is the number of occupants within the housing unit. From here we will build our model and facilitate estimation of the population values. These housing units are located within blocks, which are located within block groups, which are located in tracts. We are dealing with the aggregation of data at the tract level and so the observations are undifferentiated from each other within the tracts. For the purpose of modeling, each housing unit is assumed to be independent of other housing units.

The number of occupants within a housing unit can be as small as zero (a vacant unit) and can range up to an arbitrarily large number. Often in large occupant cases, a person is more likely dealing with a group quarter rather than a household, but it is not impossible to have a very large number of occupants within a housing unit.

As mentioned, the data are aggregated by census tracts. The data within these census tracts are assumed to be of the same common distributional form with parameters specific to the individual tract. The tract-level summary of the 2000 Census housing unit data used in this analysis is given in Appendix A.

3.2 Tract Level Marginal Model

Our strategy for picking a tract-level model was to start out with a relatively simple model that covered the salient features of the data, and then, with the help of data analysis, add complexity as needed.

3.2.1 The Initial Proposal: Zero-Independent Poisson

The first step when dealing with data is to model the data itself. In this case we tasked ourselves to construct a marginal model for the data with no covariate information or other information used for the purpose of estimation.

Our initial model was built off the idea of preserving notions of vacancy rate and persons per household. We knew our geography had within it tracts for which there were a large amount of vacants. A zero-independent Poisson model was proposed to handle this phenomenon while preserving the term for the probability of zero as solely relating to the vacancy rate. Letting t represent the index for tracts on $t = 1, ..., T$ then our model for the number of housing unit occupants has the form

$$P(X_t = x) = p_{0t} \times I(X_t = 0) + (1 - p_{0t}) \times \frac{e^{-\lambda_t}\lambda_t^x}{x!(1 - e^{-\lambda_t})} I(X_t > 0) \tag{9}$$

In our models we want to utilize a hierarchical structure so that in the cases where tracts are bereft of data we can borrow strength from other tracts to reduce error. For this reason, we let $logit(p_{t0}) \sim N(\mu_p, \sigma_p^2)$ and $log(\lambda_t) \sim N(\mu_\lambda, \sigma_\lambda^2)$. Likewise, since we are operating within a Bayesian framework we further let μ_p, μ_λ, σ_p^2, and σ_λ^2 to have vague or non-informative priors as allowed by computational reality.

Having stated our model for the data we then need to verify the model choice. Assuming that the Poisson model is suitable in general we must show that we need the separate term for zero occupants and we must show that only a term for zero occupants (and say, not for zero and one occupants separately) is needed. While there are numerous ways to invalidate a model these two ways appear to be the most obvious.

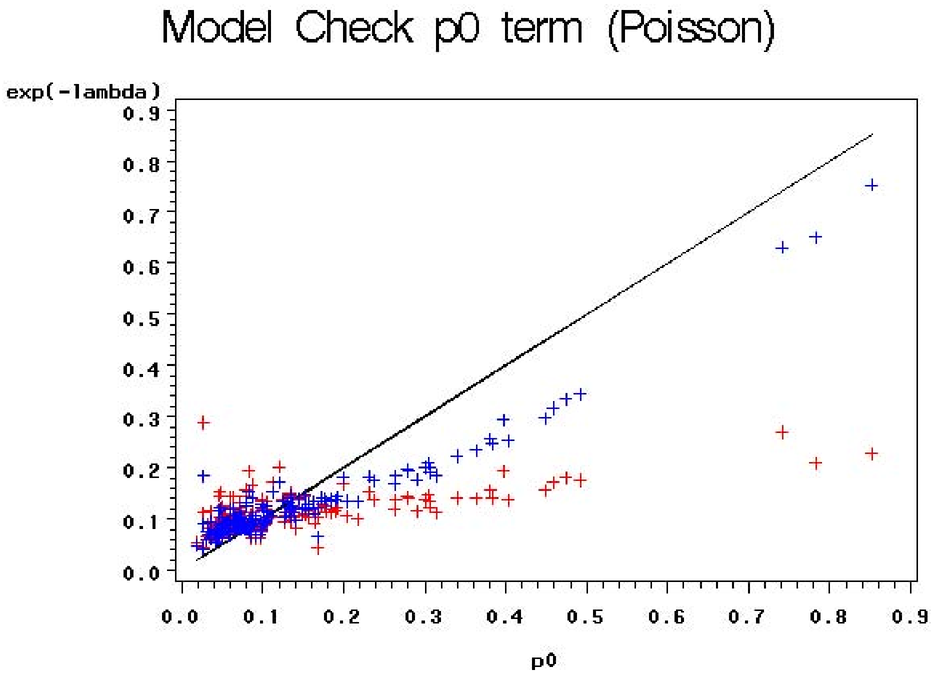
To establish the necessity of a term relating to those units with zero vacancies we recall (9). If a simple Poisson model with parameters varying by tract is sufficient to explain the data then we should be able to reduce the model. Namely, if we can show that $p_{t0} = e^{-\lambda_t}$ over the values of t then we would find that (9) reduces to the following simple Poisson model:

$$P(X_t = x_t) = e^{-\lambda_t} \times I(X_t = 0) + \frac{e^{-\lambda_t}\lambda_t^x}{x!} I(X_t > 0) \tag{10}$$

To assess if $p_{t0} = e^{-\lambda_t}$ we take the posterior mean of p_{t0} based on the model in (9) and plot it against the posterior mean of $e^{-\lambda_t}$ based on the model in (10). If we observe values close to a 45-degree line then we can simplify the model. The points obtained in this manner are shown on figure 2 in the color red. Another way would be to again estimate $e^{-\lambda_t}$ from (10) but compare it to the posterior mean of p_{t0} from (1). If the simpler model is true then we should see these values should also be nearly equal. When constructing a plot these values should also adhere to a 45-degree line. This is shown in figure 2 in the color blue.

Note that we give the same type of hierarchical structure to (10) that we did to (9).

Figure 2:



We see in figure 2 that the values do not conform to a line, but they do appear to conform to some sort of functional relationship with each other. Based on this visual inspection, we conclude that we do need the housing units with zero occupants to be modeled separately.

Next, we look at the introduction of a term in the model for those housing units occupied by one person. This distribution for this model has the following formula:

$$P(X_t = x_t) = p_{0t} \times I(X_t = 0) + p_{1t} \times I(X_t = 1) + (1 - p_{0t} - p_{1t}) \times \frac{e^{-\lambda_t}\lambda_t^{x_t}}{x_t!(1 - e^{-\lambda_t} - \lambda_t e^{-\lambda_t})} I(X_t > 1) \quad (11)$$

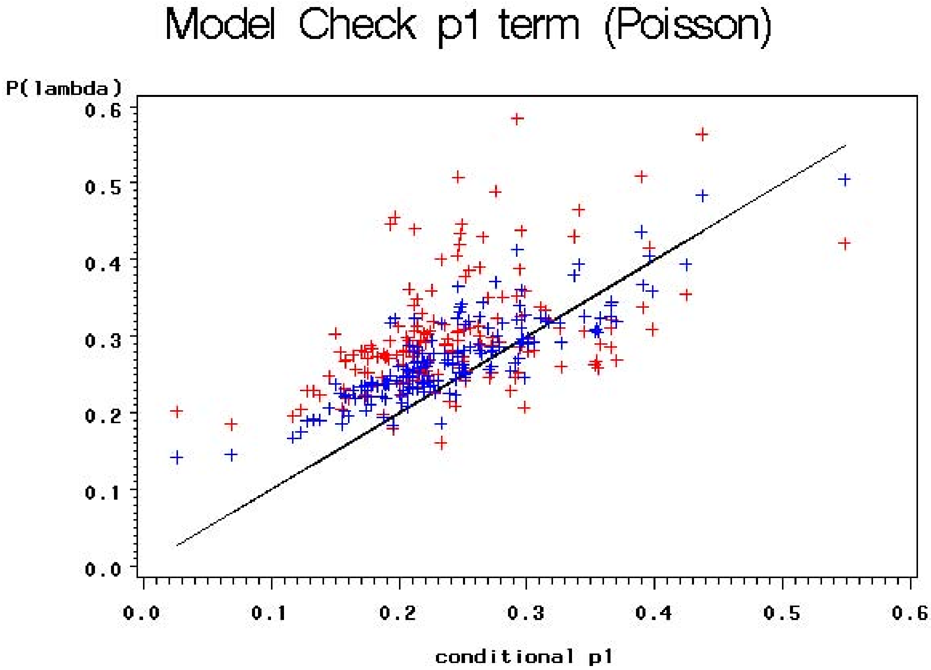
To get from (11) to (9) we would let $p_{1t} = (1 - p_{0t})\frac{\lambda_t e^{-\lambda_t}}{1-e^{-\lambda_t}}$. To simplify matters we construct plots based on $\frac{p_{1t}}{1-p_{0t}}$ and $\frac{\lambda_t e^{-\lambda_t}}{1-e^{-\lambda_t}}$ in the same manner as we did for the first plot.

In figure 3 we see some adherence to the ideal 45-degree line but not so much that it leaves us convinced of the suitability of the zero-independent Poisson model. We see that the majority of the values appear to be above this 45-degree line. We would expect that values would fall upon either side of this line. It may be that the nature of hierarchical models may be playing a role but this is uncertain. Since we are faced with these questions we look in a different direction in an attempt to conclude our model building.

3.2.2 Multinomial Modeling

Since modifying the Poisson models step by step seemed impractical and since our more complex model could still be a poor fit, we felt it would be wise to go from one end of the modeling spectrum to the other. Instead of using a parametric model, we used a semi-parametric approach. In this case we sought out the use of the multinomial distribution. Here we model the individual probabilities of observing a housing unit

Figure 3:



with a certain number of occupants. The appeal here is that we are able to model each tract with its own multinomial distribution, thereby avoiding issues of lack of fit on individual tracts or across tracts in general.

Unfortunately, the multinomial distribution does not answer all questions for us. Within the tracts there exist housing units with a large number of occupants; one housing unit within the Delmarva region has as many as 30 people. We can place a cap on the values for which we make our multinomial model, but this will have a trade-off of possibly undercounting population if predictions are made. We may also see tracts for which the number of units occupied by 6 people, for instance, is zero. While we will continue to use a hierarchical structure we may become too specific when using a multinomial distribution.

With both advantages and disadvantages in mind we can use the multinomial distribution as an exploratory data analysis tool. If we can spot any patterns then we may be able to find a different answer to our problem.

For our purposes we model zero through nine housing unit occupant counts as individual multinomial terms. Housing units of occupant size 10 and above are aggregated into a single term. The multinomial parameters are treated in a hierarchical fashion similar to 1 but using the multinomial logistic form, i.e.

$$\frac{e^{v_{tx}}}{1+\sum_{x=1}^{J} e^{v_{tx}}} = p_{tx}, v_{tx} \sim N(\mu_x, \sigma_x^2), \tag{12}$$

with vague or non-informative priors used in the appropriate fashion.

The best way to look at these probabilities is to work in a conditional fashion. Instead of displaying the probabilities, p_{tx}, Table 1 uses estimates of $q_{tx} = p_{tx}/(1 - p_{t0} - \ldots - p_{t(x-1)})$, or rather the probability of observing x occupants given that you will see $X_{th} \geq x$ occupants. This will let us see if there are any patterns in the tails of the densities. If we can find a pattern then we may be able to relate the tails through a parametric distribution.

We can glean a few features from estimated, tract-level, conditional probabilities (Table 1). One feature is

q_{t0}	q_{t1}	q_{t2}	q_{t3}	q_{t4}	q_{t5}	q_{t6}	q_{t7}	q_{t8}	q_{t9}
0.036	0.160	0.407	0.380	0.618	0.563	0.511	0.567	0.521	0.514
0.054	0.273	0.434	0.411	0.590	0.659	0.694	0.642	0.558	0.491
0.052	0.221	0.485	0.429	0.618	0.688	0.669	0.579	0.551	0.609
0.054	0.203	0.425	0.460	0.608	0.676	0.642	0.597	0.592	0.559
0.185	0.263	0.490	0.431	0.587	0.681	0.634	0.618	0.517	0.560
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.493	0.277	0.696	0.472	0.673	0.704	0.609	0.615	0.516	0.585
0.364	0.293	0.576	0.443	0.594	0.657	0.655	0.627	0.560	0.492
0.742	0.437	0.758	0.514	0.619	0.671	0.640	0.576	0.493	0.524
0.783	0.293	0.772	0.490	0.632	0.752	0.694	0.600	0.544	0.528
0.341	0.212	0.661	0.450	0.640	0.693	0.693	0.600	0.494	0.461
0.291	0.251	0.502	0.466	0.608	0.643	0.633	0.625	0.559	0.493
0.380	0.266	0.628	0.494	0.646	0.691	0.696	0.594	0.517	0.559
0.458	0.246	0.712	0.495	0.668	0.616	0.659	0.623	0.539	0.559
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.103	0.219	0.445	0.441	0.567	0.606	0.603	0.586	0.435	0.409
0.050	0.159	0.461	0.408	0.627	0.657	0.627	0.652	0.494	0.551
0.070	0.191	0.455	0.431	0.627	0.683	0.631	0.612	0.587	0.588
0.069	0.236	0.468	0.453	0.618	0.689	0.696	0.576	0.596	0.557
0.103	0.230	0.459	0.386	0.571	0.621	0.606	0.684	0.532	0.552

Table 1: Estimated Conditional Probabilities of Housing Unit Size for Selected Tracts (rows)

that we can see that the vacancy term in the model does appear to have some strong differences across tracts. We can also see that the conditional tract probabilities appear to be relatively stable for housing units of size 4 and above. This is to say that there is no clear trend but they seem to cluster around a conditional probability of 0.6. Similarly, while the first few conditional probabilities do differ in a non-obvious way, housing units with size 2 and 3 do appear to be near to each other. While this does not mean that they should be summarized by the same single parameter it may be more natural to relate terms in this conditional tail manner than to rely upon a multinomial logistic hierarchical model. Similarly, the tail conditional behavior suggests a geometric distribution model at some point.

This table also tells us that the Poisson distribution is not useful for modeling in the tail of the distribution. The reason for this is that $\frac{P(x)}{1-F(x-1)} \to 1$ as $x \to \infty$ if the data are Poisson in nature on the tails. The probabilities do not increase as they should so the Poisson distribution does not fit the data.

3.2.3 The Tail Distribution

The need of a tail distribution was implicit in the realization that the number of housing unit occupants is unbounded and so a model for larger values would be needed. We then ask ourselves, "what form does this distribution take?"

Given our observation in Table 1, we did two things: We made the tail density function a Geometric Distribution with parameter, p, as we saw that the tail probabilities are roughly the same. We made the cut-off at $k = 4$ since the probabilities appear to become constant starting there.

3.2.4 Using the Marginal Model

At this point we have modeled the marginal model for the 2000 census data upon these census tracts. We have not used any possible covariate information available to us nor have we utilized any spatial structure. Since our goal in making small area estimates is to make estimates of housing unit characteristics during non-census years, we will only have a small sample of housing units available to us and will need to rely on the best model that we can assemble in order to make predictions for the unsampled housing units. The use of good covariates as predictors can help accomplish this goal. In addition, covariates that are themselves spatially correlated may reduce or eliminate the need for a spatial model. Hence, we will want to include covariates in our model in order to fairly evaluate a spatial model. In the interest of time, the only covariate information we entertain is housing unit characteristics, at the census tract level, as measured in the 1990 Census. We chose this covariate information because it is readily available, it should provide good predictors and it should also be spatially correlated (if the 2000 housing unit characteristics are spatially correlated).

The useful thing about first developing the marginal model (i.e. a model without covariates) is that we now have a better idea of the distributional behavior of the number of occupants within a housing unit. Nevertheless, for the reasons given above, it may be useful to introduce covariate information for the tracts while also investigating the usage of spatial structure upon the tracts.

3.3 Covariate Construction

Without delving deeply into the characteristics of each tract on a variety of variables (such as race percentages, income values, etc.) we want to construct useful covariate information to reduce the uncertainty in parameter estimation. The more certain our parameter estimates the better our population estimates will be later on.

3.3.1 Using the 1990 Census

As covariates, we are constructing MLEs for the parameters in model (2) using 1990 census data. What we would like to do is simply take the 1990 data on the 2000 census tract definitions and find the estimates for the model parameters from (2). The difficulty in this is that some of the 2000 census tabulation tracts are different from the 1990 census tabulation tracts. (As an indication of the change, around 29% of the tracts defined for the 2000 census in the fifty states and D.C. exhibited a change in population of 2.5%, or more, due to a change in their boundaries from 1990*.

To overcome this difficulty we take the data from the 1990 census tracts and reorganize them on the 2000 census tracts. To do this we break apart the 1990 tracts where needed and then reaggregate them into the 2000 census tracts. We use the land area of the tracts as the mechanism to split the data. Example: Consider tracts "1" and "2." "1" is split into two parts. The first part with two-thirds of the land area remains as tract "1", the other part becomes part of tract "2." The next table demonstrates this behavior.

	1990		split			2000	
housing unit size	1	2	$1 \to 1$	$1 \to 2$	$2 \to 2$	1	2
0	10	5	6.67	3.33	5	6.67	8.33
1	15	16	10	5	16	10	21
2	12	4	8	4	4	8	8
3	5	8	3.33	1.67	8	3.33	9.67
4	1	0	0.67	0.33	0	0.67	0.33

*Calculated from public files documenting tract changes, (U.S. Census (2009)).

Once we get the 1990 census data aligned to the 2000 census tracts we can construct our covariates. Since our model terms consist of q_{t0}, q_{t1}, q_{t2}, q_{t3}, and q_{t4} from (2), by tract, we will want to make similar model estimates for these on the 1990 data. However, in order to use tract-level covariates based only on one tract at a time, we will use maximum likelihood estimates based on (2) without an additional hierarchical model (call them $\hat{q}_{[90]t0}$, $\hat{q}_{[90]t1}$, $\hat{q}_{[90]t2}$, $\hat{q}_{[90]t3}$, and $\hat{q}_{[90]t4}$). These estimates have the following form:

$$\hat{q}_{[90]ti} = \begin{cases} \frac{|\{h:X_{th}=i\}|}{|\{h:X_{th}\geq i\}|} & i \in \{0,1,2,3\} \\ \frac{|\{h:X_{th}\geq 4\}|}{|\{h:X_{th}\geq 4\}|+\sum_{i\geq 4} i\times|\{h:X_{th}=i\}|} & i = 4 \end{cases} \tag{13}$$

Since we want to treat these as model covariates and since we want to utilize a hierarchical model in the same form of the proposed tail model, we take $x_{ti} = logit(\hat{q}_{[90]ti})$ as covariates, centering them by their means to facilitate computation.

With the covariate terms in place, our non-spatial model can be generally stated as in (5) and (8):

$$logit(q_{ti}) = \alpha_i + x_{ti}\beta_i + e_{ti}, \quad i \in \{0,1,2,3,4\}$$

$$e_{ti} \sim N(0, \sigma_i^2), \quad i \in \{0,1,2,3,4\},$$

where e_{ti} and e_t represents our error components. Later this will also include our spatial hierarchical term. Ideally, if covariate information is a perfect baseline, β_i should be at or near 1. It can still be useful even if it only differs from zero as this would say that there is a relation in the parameters at the previous time point to the current time point. See Appendix B for a complete listing of the covariates used, by tract.

3.4 The Spatial Component

This section provides graphical residual maps to visually inspect how the tracts may be related and summarizes the Conditional Autoregressive model (CAR) used in our comparison.

3.4.1 Exploratory Spatial Analysis

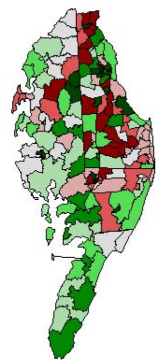
With our covariate model in place we start to consider the usage of a spatial model. One means to explore if a spatial model may be necessary is to compare the posterior means of $q_{t1}, ..., q_{t4}$ against their respective tract-level sample means that do not rely on model assumptions.

We use the logistic scale to keep the analysis on a linear basis as we deal with tracts that have very different values, especially on the vacancy term.

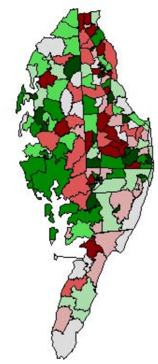
The chloropleth plots listed in figures 4(a) - 4(e) are constructed from the differences of the model and true census values for each of the parameters and they are divided by the census tracts. These plots provide the residual error that was not predicted from the non-spatial hierarchical model. As a reference, a map of uniform random numbers scaled the same as the residual maps is included in figure 4(f). From these plots it is difficult to discern an obvious spatial signal.

The plots themselves do show some apparent clustering of positive and negative differences but some clustering is expected just by the nature of random variation.

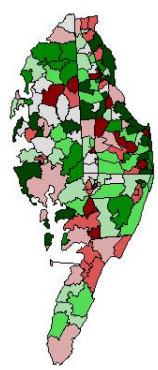
Figure 4: Map of Tract-level Residuals



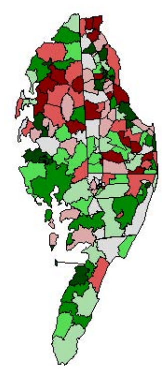
(a) q_{i0} terms



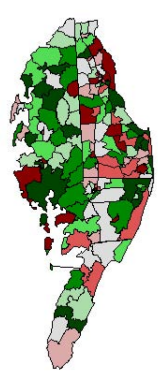
(b) q_{i1} terms



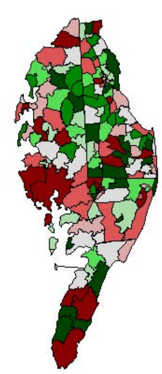
(c) q_{i2} terms



(d) q_{i3} terms



(e) q_{i4} terms



(f) uncorrelated random terms

3.4.2 Using the CAR Model

The Conditional Autoregressive model is a model which can allow us to utilize neighboring tracts and some properties of those tracts to investigate the existence of spatial structure.

There are two equivalent statements for the definition of the CAR model: the full model and the conditional model. We apply a CAR model to each of the five components, $q_{t0}, ..., q_{t4}$, separately. For component, i, $i \in \{0, 1, 2, 3, 4\}$, define, generically: $s_t = e_{ti}$. For component i, the full model and conditional model specifications are, respectively:

$$\mathbf{s} \sim N(\mu, \nu(\mathbf{I} - \gamma\mathbf{C})^{-1}\mathbf{M})$$

$$s_i|\mathbf{s}_{(-i)} \sim N(\mu_i + \gamma \sum_{j=1}^{T} C_{ij}(s_j - \mu_j), \nu M_{ii}),$$

where $C_{ii} = 0$ and $\mathbf{M}$ is a diagonal matrix with positive diagonal elements. It must also be noted that for $(\mathbf{I} - \gamma\mathbf{C})^{-1}\mathbf{M}$ to be a valid covariance structure, $C_{ij}M_{jj} = C_{ji}M_{ii}$ must be true and $|\gamma|$ must be restricted (see GeoBUGS (2004) for details).

The conditional form of the model is more intuitive as we see that it behaves like a linear model. For our purposes $C_{i,j}$ and M_{ii} are fixed, leaving γ as the sole "covariate" term of the conditional model. γ is our proxy for spatial dependence. If we can show that it is different from zero then we have evidence of spatial dependence within the model. For the purpose of Bayesian inference we let ν have a vague prior and γ has a uniform prior between its bounds.

There are many choices for C_{ij} and M_{ii} based on the information available to us. Our most basic choice is to let $C_{ij} = 1$ for neighboring pairs of tracts, $C_{ij} = 0$ for disconnected tracts, and let $M_{ii} = 1$. There does exist an inter-play of the choices of C and M. Changing M will increase or decrease the initial (no data) variation in the parameters and changing C can shift the parameter estimates depending on the values of the other tracts as shown through the γ term. As a generalization we can write a valid choice of C and M in the following manner:

$$C_{ij} = \frac{W_{ij}}{\sum_{j \in B_i} W_{ij}} \tag{14}$$

$$M_{ii} \propto \frac{1}{\sum_{j \in B_i} W_{ij}} \tag{15}$$

W_{ij} needs to be symmetric but can be any weight that one may want to apply. As defined following equation (6), B_i is the set of neighbors of i. In our case we sought to use perimeter distance information and centroid values but had difficulty in obtaining these values due to various challenges inherent in the boundary file we were using. We were able to define neighboring pairs of tracts that shared a boundary, including only a corner point. So, we used the corresponding basic specification: $C_{ij} = 1$ for these neighboring pairs of tracts, $C_{ij} = 0$ for disconnected tracts, and let $M_{ii} = 1$ throughout.

3.5 Evaluation of CAR model using complete 2000 Census Data

We choose one spatial model to evaluate. This is the aforementioned "basic" CAR model with $C_{ij} = 1$ if the tracts are neighbors, $C_{ij} = 0$ otherwise, and $M_{ii} = 1$ (see Appendix C for a listing of the neighbors). This is perhaps the simplest CAR model in use. Estimates using GeoBUGS (2004) can be readily made given the data and a map boundary file, where tracts with common boundary are neighbors. We obtained estimates of model parameters from both the hierarchical model (with covariates, referenced as the "non-spatial model") and the spatial model (also with covariates) using the full 2000 census data. We obtained 10,000 samples from 5 MCMC chains after a burn in phase of 4,000 samples. Summaries of the posterior parameter

			quantiles		
	Mean	Std. Dev.	2.5%	Med	97.5%
α_0	-2.146	0.03462	-2.212	-2.146	-2.081
α_1	-1.186	0.01708	-1.221	-1.186	-1.153
α_2	-0.05643	0.01571	-0.08754	-0.05622	-0.02543
α_3	-0.2408	0.01534	-0.2702	-0.2411	-0.21
α_4	0.5366	0.01652	0.5407	0.5363	0.5697
β_0	1.806	0.06574	1.673	1.805	1.947
β_1	1.072	0.05039	0.974	1.071	1.174
β_2	0.9997	0.04887	0.9008	1.001	1.094
β_3	0.7193	0.09078	0.5461	0.7175	0.8982
β_4	0.6925	0.08373	0.5293	0.693	0.8557
σ_0	0.4313	0.02524	0.3851	0.4302	0.4838
σ_1	0.207	0.01238	0.1843	0.2064	0.2328
σ_2	0.1912	0.01135	0.1704	0.1906	0.2149
σ_3	0.1744	0.01074	0.1547	0.174	0.1968
σ_4	0.1929	0.0121	0.1708	0.1923	0.2181

Table 2: Posterior parameter summaries from the non-spatial hierarchical model

distributions for the non-spatial hierarchical model are presented in table 2 and those for the spatial model are presented in table 3.

There are a few features to see here as it relates to the census data on the Delmarva region. With regards to spatial signal we see some evidence within the credible interval information that there is some spatial dependence for the vacancy model term (i.e., the 95% quantile bounds for γ_0 excludes zero) but for the other terms there do not appear to be enough evidence to say that there is spatial dependence (i.e., the 95% quantile bounds for the $\gamma_1, ..., \gamma_4$ each includes zero). (Note that a Bonferroni adjusted simultaneous interval for all five parameters amounts to only a 75% simultaneous interval, and, hence, this observation may have no practical significance.) Regardless of the magnitude of the posterior probability γ_0 near the value 0, its non-zero values may have an effect on the posterior inference for other parameters. (In other words, if the spatial model is better, one would expect that the posterior distribution of the other parameters of the model would become more precise. However, examining the corresponding parameters in tables 2 and 3, there does not appear to be a great reduction.) We do have a lot of data but only upon the 165 tracts dealing with the tract-to-tract characteristics. With more tracts one may see a finer estimate, but doing so would move us away from assessing the Delmarva region and towards analyzing a larger geographic region.

In both models we see that the β terms all suggest positive association between the covariate values from the 1990 census and the values for the current census. This is evidenced by observing that the credible intervals exclude 0 for each β.

4 Repeated Sampling Evaluation of Tract-level Small Area Estimates

To this point we have demonstrated our model and its application on the census data, and began evaluating the posterior distribution of the spatial correlation parameter. Each of these have their utility but our main desire is to investigate the usage of this model upon a small sample taken during an intercensal year. From the survey we seek to estimate such values as total population, the count of vacant housing units, the vacancy rate, and the person per household. We want to get estimates for these values at the census tract

			quantiles		
	Mean	Std. Dev.	2.5%	Med	97.5%
α_0	-2.129	0.03915	-2.211	-2.13	-2.055
α_1	-1.182	0.01495	-1.21	-1.182	-1.151
α_2	-0.05893	0.01574	-0.08985	-0.05868	-0.02836
α_3	-0.2388	0.01509	-0.2685	-0.2391	-0.2089
α_4	0.5354	0.01605	0.5036	0.5352	0.5668
β_0	1.754	0.06485	1.623	1.766	1.858
β_1	1.063	0.03425	0.9989	1.063	1.133
β_2	0.9783	0.03961	0.8999	0.9788	1.052
β_3	0.7031	0.07265	0.5643	0.7025	0.8463
β_4	0.6856	0.0668	0.5556	0.6847	0.8219
σ_0	0.414	0.02493	0.3686	0.413	0.4659
σ_1	0.2057	0.01218	0.1834	0.2052	0.2311
σ_2	0.1897	0.01125	0.1692	0.1892	0.2131
σ_3	0.1727	0.01064	0.1531	0.1722	0.1948
σ_4	0.1918	0.01189	0.1699	0.1913	0.2165
γ_0	0.09526	0.02938	0.02854	0.09893	0.1413
γ_1	0.02425	0.04661	-0.0747	0.02706	0.1068
γ_2	0.03875	0.0455	-0.05881	0.04217	0.1171
γ_3	0.05026	0.04351	-0.04479	0.05388	0.1237
γ_4	0.03455	0.04682	-0.06533	0.03796	0.1156

Table 3: Posterior parameter summaries from the spatial model

Finite Population Parameter	Tract Specific	Over all Tracts
Population	$X_t = \sum_{h=1}^{H_t} X_{th}$	$X = \sum_t X_t$
Vacancy Count	$V_t = \sum_{h=1}^{H_t} I_{[X_{th}=0]}$	$V = \sum_t V_t$
Vacancy Rate	$VR_t = V_t/H_t$	$VR = V/H$
Persons Per Housing Unit	$PPH_t = X_t/(H_t - V_t)$	$PPH = X/(H - V)$

Table 4: Definition of Finite Population Parameters

level. We will also look at estimates for the entire Delmarva region, obtained by summing up the tract-level components that make up the regional population parameters of interest.

4.1 Finite Population Parameters of Interest

Let H_t be the total number of housing units in tract t. Using the definition of household size from (2.1), the finite population parameters of interest are defined in table 4.

Following the Bayesian paradigm, small area estimates will be obtained from the posterior predictive distribution of the finite population parameters (e.g., Scott and Smith (1969)). A posteriori, only the unsampled observations are unknown. Due to the independence assumptions made, their predictive distribution can be obtained from their conditional distribution as specified in (1), (2) and (3) coupled with the posterior distribution of the underlying parameters. For estimation, the posterior mean will be used as the small area estimate since it minimizes a quadratic loss function and the posterior variance will be used as the corresponding measure of uncertainty. For example, the posterior mean of the population count is:

$$E(X_t|\text{sample}) = \sum_{h\in s_t} X_{th} + \sum_{h\in \overline{s}_t} E(X_{th}|\text{sample}), \tag{16}$$

where, s_t denotes the set of sampled housing units in tract t, and $\overline{s}_t$ denotes the set of housing units in tract t that were not in sample.

Posterior means of vacancy counts and vacancy rates can be made similarly, although vacancy rates are not linear functions of the predictive outcomes.

In a production setting, given a sample, we would numerically estimate the posterior mean and posterior variance of the finite population of interest and be finished. However, we want to evaluate how this whole procedure behaves over repeated samples. Even though our development is completely Bayesian, our procedure should still have good repeated sampling properties for it to be of value to a wide variety of users.

Since we have 2000 census data at our disposal, the easiest method would be to draw a sample from the 2000 census on the Delmarva region at a rate similar to that for the ACS. The ACS has a relatively simple sample design consisting of a systematic sample of housing units within four size-based strata defined nationwide (see U.S. Census Bureau (2003)); the strata are defined at the block group level. Replicating the ACS strata sample within Delmarva would have required obtaining block-level design files which would have delayed this demonstration project. In our case we simplified the sample design by eliminating the ACS strata, instead drawing simple random sample without replacement (SRSWOR) of housing units in the census based on a 2.4% sampling rate. From this we can get a grasp on the accuracy of the estimates by comparing them to the true values given by the census data.

Our goal is to evaluate the hierarchical model proposed earlier against the hierarchical spatial CAR model and determine which may perform better.

4.2 Repeated Sampling Study

Our repeated sampling consists of 400 samples of Delmarva 2000 census data. The data were sampled at a rate of 2.4% as a SRSWOR across the entire region as mentioned earlier. We did not attempt to sample on a tract level basis. Our hierarchical model and our hierarchical spatial model were then applied to these samples. These models were executed through the WinBUGS (2003) software and the GeoBUGS (2004) extension.

Within this we sought out model comparisons of bias, mean square error, and credible interval coverage rates at Delmarva level and tract level. Many of these were obtained; for the sake of brevity, not all are included here.

The bias and MSE considered in our evaluation are given in the following description. Let θ represent the true value for our quantity of interest at the Delmarva level (e.g., the total household population of the Delmarva Peninsula). Let θ_t represent the true value for our quantity of interest at the tract level (e.g. the actual vacancy rate in tract, t). Let $\widehat{\theta}_s$ represent the Delmarva level estimate for the s-th repeated sample and let $\widehat{\theta}_{s,t}$ represents the tract level estimate for the s-th repeated sample from the t-th tract. Let M be the number of repeated samples; in our case $M = 400$.

In formal terms we have the bias which takes the form of $B = (\frac{1}{M}\sum_{s=1}^{M}\widehat{\theta}_s) - \theta$ whether at the tract level or Delmarva level.

Mean Square Error is $MSE = \frac{1}{M}\sum_{s=1}^{M}(\widehat{\theta}_s - \theta)^2$ whether at the tract level or Delmarva level.
Mean Square Error aggregated across tracts is $MSE_{agg} = \frac{1}{M}\sum_{s=1}^{M}\sum_{t=1}^{T}(\widehat{\theta}_{s,t} - \theta_t)^2$. Since this is an aggre-

	Non-Spatial	Spatial	2.5% difference	97.5% difference	design-based
Bias	226.266	252.3122	-1125.878	973.904	0*
MSE (Population)	28,168,364	28,111,009	-11,669,549	11,501,894	28,384,011
MSE (Tract Aggregation)	11,249,222	11,271,673	-641,635.1	538,033.5	28,924,298

Table 5: Repeated sampling bias and MSE for estimates of population

	Non-Spatial	Spatial
50%	51.50%	50.25%
80%	82.25%	81.50%
90%	91.25%	90.75%
95%	94.75%	94.75%

Table 6: Percent of time coverage intervals actually cover the true population

gation across tracts, this is only a Delmarva level assessment.

We are also interested in the rates of credible interval coverage, or $\frac{1}{M}\sum_{s=1}^{M} I(\theta_s \in (\widehat{\theta}_{s[\alpha/2]}, \widehat{\theta}_{s[1-\alpha/2]}))$. Likewise we can consider this on an individual tract basis or at the Delmarva level.

Since we are operating with a non-linear model in the Bayesian paradigm, obtaining estimates is not so straight-forward. Due to time constraints we let our model be a single-chain from an MCMC of 1000 burn-in iterations followed by 1000 iterations of samples from the density. Longer runs of the MCMC chain would result in more precise estimates of posterior parameters.

4.2.1 Results of the Repeated Sampling Study

To start the analysis we looked at the performance of the bias and mean square error at the Delmarva level for the population estimate. At this point a design-based (i.e. non-model based) estimate for the population in each tract was introduced for the purpose of acting as a baseline with which to compare the hierarchical and spatial model estimates. By running GeoBugs for a long number of iterations for one sample for both spatial and non-spatial models, we observed that the numerical bias caused by running GeoBugs for only 400 iterations did not appear to be too influential as the value is smaller than one person (not shown). Bayes estimates tend to be biased estimates but this suggests that the performance in terms of bias is good. Looking at the MSE as it relates to the overall population, the spatial model appears to perform slightly better than the non-spatial model. Table 5 appears to show that the aggregate tract MSE for the spatial model is better than that for the non-spatial model. This table also provides 95% confidence bounds for the differences between the non-spatial and spatial estimates. Based on these bounds, there is no statistical difference in the bias or MSE between the spatial and non-spatial estimates of population size. Lastly, table 5 documents that, at the tract level, both the spatial and non-spatial models provide estimates with much smaller MSE than the design based estimates, demonstrating the power of "borrowing strength" using small area estimation methodology.

Next, table 6 estimates the percentage of time each designated coverage interval of total population actually covers the true value. Looking at the associated credible intervals for both models at the 50%, 80%, 90%, and 95% levels we see that there does not appear to be any clear evidence of over-coverage or under-coverage.

*The bias under this model is known to be zero.

	Non-Spatial	Spatial	2.5% difference	97.5% difference
Vacancy Count Bias	29.68679	26.21703	-216.184	255.030
Vacancy Count MSE (Pop.)	1563084	1559765	-603909.3	643161.4
Vacancy Count MSE (Tract Agg.)	938303.8	936672.0	-71678.98	83920.31
Vacancy Rate Bias	$2.163 * 10^{-7}$	$1.911 * 10^{-7}$	-0.0006303365	0.0007436014
Vacancy Rate MSE (Pop.)	$1.329 * 10^{-5}$	$1.326 * 10^{-5}$	-5.134162e-06	5.467865e-06
Vacancy Rate MSE (Tract Agg.)	0.1707873	0.1692201	-0.006998501	0.010901630
PPH Bias	$1.139 * 10^{-3}$	$1.121 * 10^{-3}$	-0.003438485	0.003088989
PPH MSE (Pop.)	$2.706 * 10^{-4}$	$2.700 * 10^{-4}$	-0.0001365973	0.0001182237
PPH MSE (Tract Agg.)	2.003925	2.017699	-0.1259202	0.0937945

Table 7: Repeated sampling bias and variance for estimates of vacancy and PPH

Turning our attention to the vacancy count, vacancy rate, and persons per household statistics, in table 7, it would appear that for bias, MSE at Delmarva level, and MSE aggregated by tract, the spatial model appears to perform better but in a way that is not significant.

Figure 5 shows the coverage rates, by tract, of the credible intervals for the population. We saw some interesting phenomena in the plots where the trend was for over-coverage on many of the tracts but there were some severe outliers on the under-coverage side of the plots. Our plots exhibited a stronger leftward skew than what we would otherwise see from natural variation.

Lastly, we look at the coefficient of variation, σ/μ by tract for the population estimate by tract on our two models, in figures 6(a) and 6(b), and the design-based estimate in figure 6(c). The two models produce nearly the same result and they both performed better than the design-based estimate, again underlying the decrease in variation against the best traditional estimate available using this data.

5 Remarks

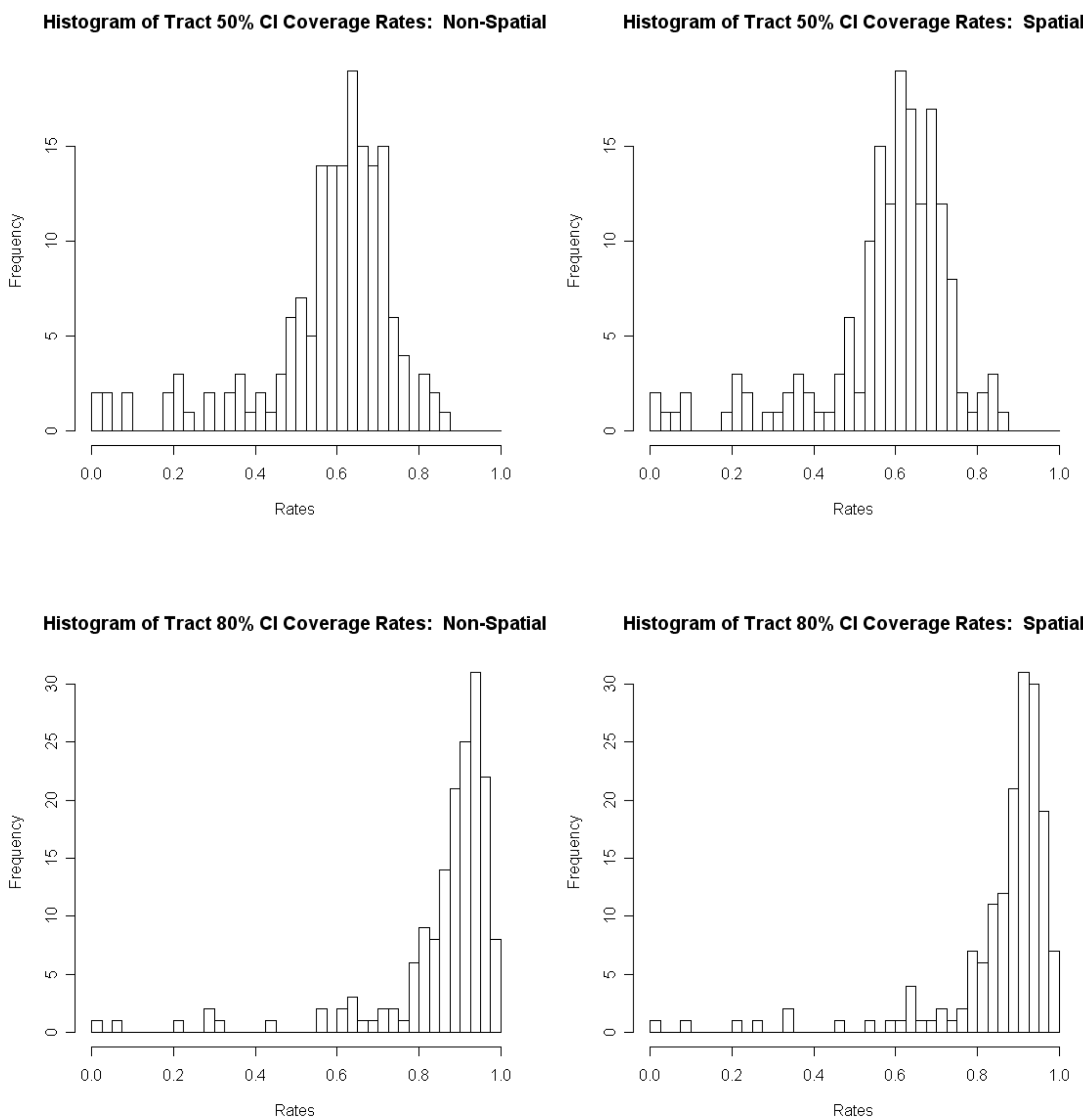
In the project there were a few issues and concerns to address and obstacles that we needed to overcome. Naturally, as with any project, more time would be have been useful to answer other questions, but it is also useful to outline the limitations of the project.

5.1 Difficulties

To complete this project we utilized SAS, WinBUGS, R, and GeoBUGS, which is an internal branch of WinBUGS. The usage of WinBUGS and GeoBUGS proved to the most problematic aspect of our project. Naturally, when using MCMC methods, time constraints may begin to impose themselves. In our case, as we worked with more and more tracts on more and more geographies we ran into serious computation issues which was natural considering the stark increase of the number of tracts. Similarly, when applying the spatial model to the larger geographies our desktop computer (a Dell Optiplex GX280 running Windows SP professional) used for computation could not allocate enough memory.

We also were not able to use the GeoBUGS option to use a variety of spatial adjacency information in our models. Some of the models we tried that satisfied the necessary variance constraint would not run for us in GeoBUGS and vice versa. This was in relation to the CAR spatial modeling for C_{ij} and M_{ii} not explicitly equal to some constant. In some cases the software would accept invalid structures and in other cases it would reject valid structures on the basis that the covariance matrix was not positive definite. This made investigating ways in which neighboring tracts may influence each other difficult. The investigation of modeling C and M was an initial item of interest when we started our investigation. Alternative avenues

Figure 5: Distribution of Tract-level Estimated Coverage Rates



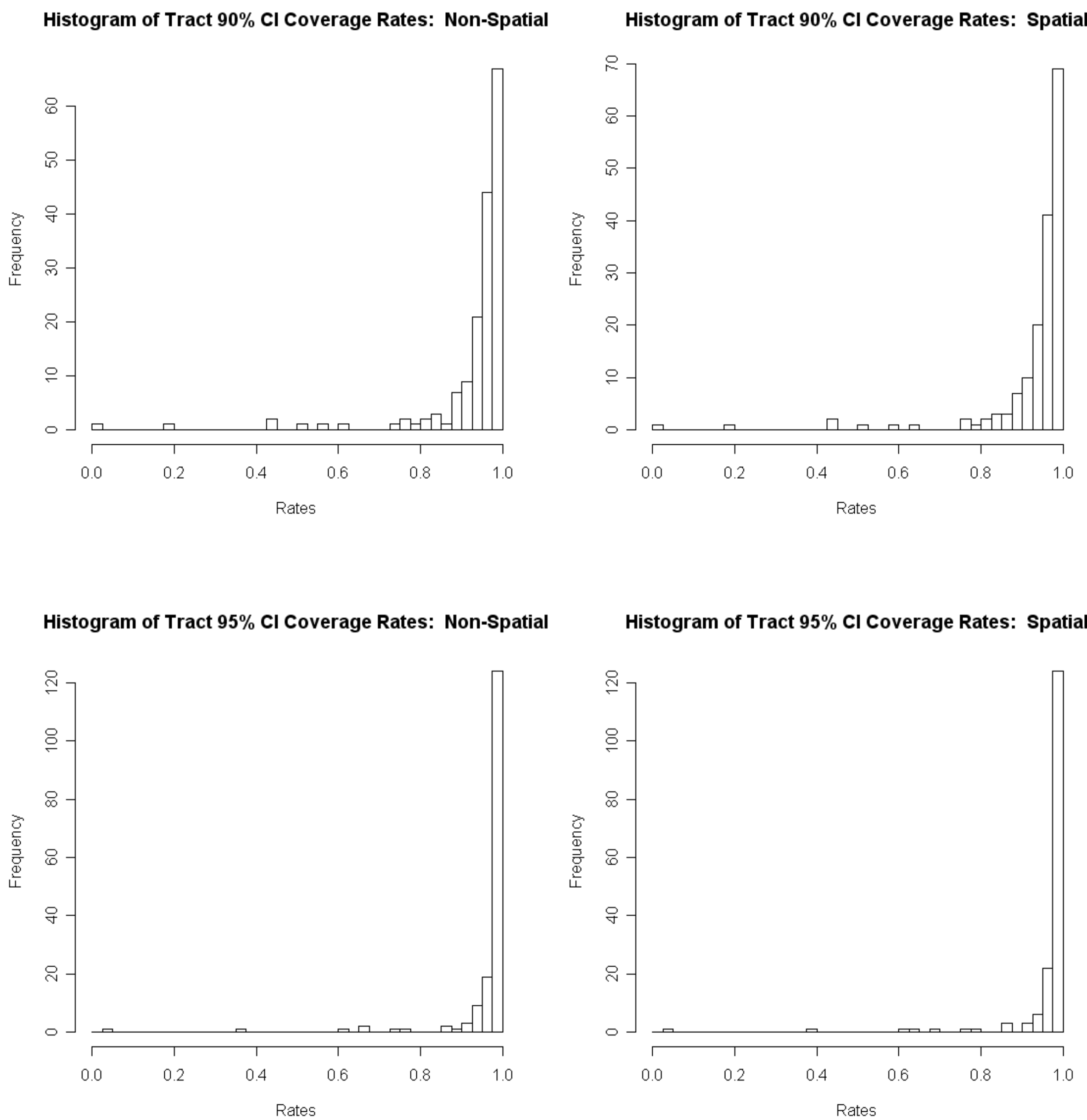
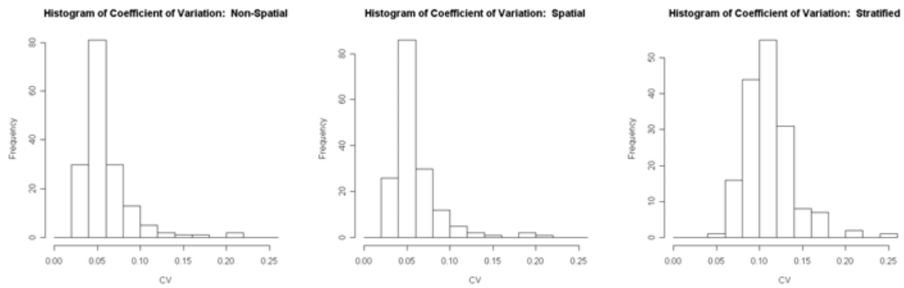


Figure 5: Distribution of Tract-level Estimated Coverage Rates (Cont.)

Figure 6: Distribution of Tract-level CVs



(a) **(b)** **(c)**

may need to be pursued if one is to investigate using the CAR model within this context.

Another issue which came up during our investigation was the use of the publicly available, generalized ARC/INFO Ungenerate (ASCII) Cartographic Boundary Files relating to the 2000 census data. GeoBUGS has a tool which automatically detects neighbors when imported into a proper format. Our issue is not with this format but that the boundaries themselves did not suit our needs (i.e., identifying tracts with shared perimeters, non-identification of neighbors). While the boundaries for tracts match each other within state we found instances where along state boundaries the points used to define the state line for those tracts would differ resulting in a non-identified neighbor. This issue also meant that we could not accurately (and consistently) define objects like shared tract perimeter which we desired to use in proposals for CAR model terms. We learned subsequently that these problems arose because the boundary files we used are a generalized extract from the U.S. Census Bureau's TIGER database and were subject to line simplification/smoothing, coordinate reduction, and small polygons were eliminated when the combination of geographic codes existed elsewhere. Thankfully, GeoBUGS has a mechanism to manually augment and alter the identification of neighbors allowing us to add and subtract neighbors as needed and to finish this project. If this project were done over multiple states we would have difficulties in adding and subtracting, both because the number of revisions could be large and also because the system is graphically based, requiring tract boundaries to be visually identifiable. Although requiring an extra processing step and more software, tract-level boundary files with matching neighbor boundary points are available and will be used in any future work. Instead of using the ASCII tract-level boundaries, with their inherent problems, non-generalized tract boundaries are available and can be freely downloaded through the National Historic Geographic Information System[†]. These shapefiles can then be translated into ASCII files using commercial GIS software and then used as input to GeoBugs. For the complete overview of the Cartographic Boundary File generalization process see U.S. Census Bureau (December 31, 2008).

5.2 Data Quirks

Along the way we came across some interesting features in the data and the tract level geography. Within the Delmarva region one tract was nested within the space of another tract. This tract contained no housing units but did, on further investigation using an overlay of census tracts on a Google map from the Internet, contain a prison. We came to find out later that prisons often are assigned their own tracts but this is not something that is done on a consistent basis. There may be a rule governing such features but we are not aware of it. Since this prison did not contain housing units in either census collection (and more specifically in the 2000 collection), it does not contribute to the residential population and thus we can omit this tract

[†] Minnesota Population Center, University of Minnesota, http://nhgis.org/

from the analysis.

Another quirk revealed itself when testing out the model on other geographies to see what would happen. We found a situation in which a tract contained no persons in the 1990 census but did so in the 2000 census. This would be a cause for modeling concern as this type of situation would have no available covariates for usage in the model. We believe remedies could be proposed to alleviate such phenomena, but we were not going to concern ourselves with this behavior as we were using the Delmarva region as our dataset.

5.3 Remarks

We have picked tract level as our small area of interest, and arbitrarily used tracts as the basic component of spatial correlation. Although this use of tracts will capture spatial variation due to larger or different geographic scales, the scale at which the outcomes of interest become homogeneous has not been investigated. An approach using the data to inform the appropriate geographic scale of the data, such as in Louie and Kolaczyk (2006), may be of use. Another way to approach the problem would be to operate on the block or block-group level. It may be that these objects will have stronger spatial signal since neighboring blocks or block groups are geographically closer than neighboring tracts. On the other hand by separating into smaller units the inferential power on any one unit will be smaller and thus may not necessarily lead to a more precise estimate for spatial correlation. Working with smaller units will also pose a challenge in computation, similar to the problems mentioned earlier, and may pose its own challenges when dealing with geographic shape. Also, we have evaluated estimates based on MSE. Another evaluation which may be important in some applications is how well the true spatial correlation is maintained. For this latter concern, a spatial model is likely to be of more importance.

5.4 Conclusions

With regards to the housing unit model, a model for housing unit occupant size based on non-parametric terms for small values and a parametric term for larger values appears to be the best option. We saw flaws in the usage of simple parametric models, specifically the Poisson distribution. There was plenty of evidence to suggest that the tracts behaved differently and should not be aggregated into a single summary parameter for all the tracts. This was most evident in modeling the housing unit vacancies as evidence by their associated estimated variance shown in tables 2 and 3. This was also evidenced by the consistency of conditional probabilities for one occupant, two occupants, and three occupants. The tail behavior, as evidenced by the appearance of consistently equal probability values, led to using a geometric distribution on those tails as a better model. However, the slow decrease observed in the estimated conditional tail probabilities, apparent in table 1, suggests that more modeling, or at least formal testing, may yield better results.

From the analysis, the case for spatial dependence on the Delmarva geography as illustrated through the CAR model is plausible but not a certainty. If spatial signal does exist then it appears to exist through the vacancy term. As the vacancy term also has more variability associated with it, it just might be that the covariates used may have not been able to remove spatial dependence - a conjecture which could be further investigated. In our experience with this project, the use of GeoBugs provided excellent support when using the simplest CAR model denoting spatial correlation based only on tract boundary contiguity. However, implementing other CAR models in GeoBugs presented problems. Hence, our investigation of spatial behavior is not exhaustive, and there may yet be a manner in which a healthy spatial signal is revealed. Further modeling may require writing specialized software or, at a minimum, further ascertaining the strengths and weaknesses of GeoBugs.

The repeated sampling study does not assure that using a hierarchical model with a spatial component is better than using only a hierarchical model. With more samples we may be able to establish that the spatial model does perform better in estimation than the non-spatial model but we were unable to discern any advantage. It is also important to recall that we used only one type of spatial model, a CAR model postulated independently for each household size component, and we used only the simplest spatial correlation structure (tracts that share a county or not) with our CAR model. However, we suspect that if spatial correlation had been an extremely important component of household size in the Delmarva Peninsula, our basic model would have shown better results.

6 References

Bell, William R. (2008). Examining Sensitivity of Small Area Inferences to Uncertainty About Sampling Error Variances ASA Proceedings of the Section on Survey Research Methods,

Chand, Nanak and Charles H. Alexander. (1999). Indirect estimation based on administrative records and the American Community Survey ASA Proceedings of the Section on Survey Research Methods, 871-876

Chand, Nanak and Charles H. Alexander. (1996). Small area estimation with administrative records and continuous measurement ASA Proceedings of the Section on Survey Research Methods, 870-875

Chand, Nanak and Charles H. Alexander. (1995). Indirect estimation of rates and proportions for small areas with continuous measurement ASA Proceedings of the Section on Survey Research Methods, 549-554

Chand, Nanak and Donald Malec. (2001). Small Area Estimates from the American Community Survey Using a Housing Unit Model, proceedings of the 2001 Federal Committee on Statistical Methodology conference http://www.fcsm.gov/events/papers2001.html

Louie, Mary M. and Eric D. Kolaczyk. (2006). A multiscale method for disease mapping in spatial epidemiology, Statistics in Medicine, 25, 8, 1287-1306

Malec, Donald. (2005). Small area estimation from the American community survey using a hierarchical logistic model of persons and housing units Journal of Official Statistics, 21, 411-432

Ghosh, Malay, Kannan Natarajan, T. W. F Stroud and Bradley P. Carlin. (1998). Generalized linear models for small-area estimation Journal of the American Statistical Association, 93, 273-282

GeoBUGS User Manual. (2004). http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs12manual.pdf

Rao, J. N. K. (2003). *Small Area Estimation*, John Wiley.

Scott, Alastair and T. M. F. Smith. (1969). Estimation in Multi-Stage Surveys *Journal of the American Statistical Association*, 64, 327, 830-840

U.S. Census Bureau. (February 3, 2009). Census Tract Relationship Files
http://www.census.gov/geo/www/relate/rel_tract.html

U.S. Census Bureau. (December 31, 2008). Cartographic Boundary Files: Scale, Generalization, and Limitations of the Cartographic Boundary Files
http://www.census.gov/geo/www/cob/scale.html

U.S. Census Bureau. (2006). The American Community Survey Technical Paper 67: Design and Methodology, http://www.census.gov/acs/www/Downloads/tp67.pdf

U.S. Census Bureau. (2005). ARC/INFO Generate (ASCII) Metadata Cartographic Boundary Files,
http://www.census.gov/geo/www/cob/ascii_info.html

WinBUGS User Manual. (2004). http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf

APPENDIX

A Counts of Housing Unit Size by Tract (Census 2000)

state	county	tract	Household count by size of household																		
			0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	19	30
10	001	040100	68	294	630	346	358	120	48	30	12	6	3	1	2	0	0	0	0	0	0
10	001	040201	77	375	431	230	196	92	35	10	2	0	1	0	0	0	0	0	0	0	0
10	001	040202	105	422	723	328	271	118	35	9	4	3	0	0	0	0	0	0	0	0	0
10	001	040203	65	234	391	247	173	77	23	8	4	1	0	0	0	0	0	0	0	0	0
10	001	040400	107	125	169	75	57	31	8	4	0	1	0	0	0	0	0	0	0	0	0
10	001	040500	304	785	1213	597	425	181	62	21	7	4	0	0	0	0	0	0	0	0	0
10	001	040600	31	134	93	47	34	12	4	2	0	0	0	0	0	0	0	0	0	0	0
10	001	040700	82	365	539	330	254	107	41	11	1	1	1	0	0	0	1	0	0	0	0
10	001	040800	59	442	384	187	109	44	19	9	2	1	0	1	1	0	0	0	0	0	0
10	001	040900	27	577	279	98	61	20	4	0	0	0	0	0	0	0	0	0	0	0	0
10	001	041000	146	563	650	324	193	101	43	11	7	1	2	0	0	0	0	0	0	0	0
10	001	041100	213	15	257	338	336	79	26	3	0	1	0	0	0	0	0	0	0	0	0
10	001	041200	96	336	456	251	155	93	36	7	4	2	1	0	0	0	0	0	0	0	0
10	001	041300	81	352	313	132	97	36	14	6	2	0	0	0	0	0	0	0	0	0	0
10	001	041400	144	442	356	206	136	57	25	10	4	0	0	2	0	0	0	0	0	0	0
10	001	041500	70	265	524	297	181	99	40	10	2	0	0	0	0	0	0	0	0	0	0
10	001	041600	43	196	354	157	120	44	28	3	0	1	0	0	0	2	0	0	0	0	0
10	001	041701	94	364	536	273	196	84	36	9	4	1	1	0	0	0	0	0	0	0	0
10	001	041702	54	155	367	239	237	102	35	5	2	2	0	0	0	0	0	0	0	0	0
10	001	041801	129	609	934	562	442	181	59	31	10	4	4	2	4	0	0	0	0	0	0
10	001	041802	41	167	295	165	155	57	17	17	9	7	5	2	0	1	1	0	0	0	0
10	001	041900	83	306	612	310	308	123	36	13	12	3	3	2	4	0	0	0	0	0	0
10	001	042000	63	181	368	187	169	90	34	14	4	3	3	0	0	0	0	0	0	0	0
10	001	042100	62	253	468	241	173	63	22	4	4	6	0	2	1	0	0	0	1	0	0
10	001	042201	220	317	640	433	339	222	73	27	6	1	2	0	0	1	0	0	0	0	0
10	001	042202	91	410	700	427	399	155	51	19	6	2	1	0	0	0	0	0	0	0	0
10	001	042400	183	233	400	173	158	55	18	5	0	0	0	0	0	0	0	0	0	0	0
10	001	042500	93	435	347	192	149	65	20	10	2	0	0	1	1	0	0	0	0	0	0
10	001	042600	22	194	331	155	118	39	15	2	2	0	0	0	0	0	0	0	0	0	0
10	001	042700	41	114	151	73	70	24	11	4	0	2	0	0	1	0	0	0	0	0	0
10	001	042800	113	374	738	404	376	107	43	11	5	1	0	0	0	0	0	0	0	0	0
10	001	042900	65	214	448	246	202	92	31	13	4	1	0	0	0	0	0	0	0	0	0
10	001	043000	128	439	583	292	237	114	50	14	7	3	1	0	0	0	0	0	0	0	0
10	001	043100	40	154	328	165	134	71	15	8	6	3	0	1	0	0	0	0	0	0	0
10	003	016601	80	229	570	382	450	184	40	19	5	3	1	2	0	0	0	0	0	0	0
10	003	016602	36	87	398	272	388	167	36	11	3	3	0	0	1	0	0	0	0	0	0
10	003	016603	22	142	367	242	314	135	32	14	1	2	0	0	0	0	0	0	0	0	0
10	003	016604	153	402	562	361	323	133	35	12	5	5	3	1	0	0	0	0	0	0	0
10	003	016801	55	150	379	205	207	72	34	6	1	2	0	0	0	0	0	0	0	0	0
10	003	016802	43	104	296	159	183	66	29	3	4	2	0	0	0	0	0	0	0	0	0
10	003	016901	27	124	288	172	138	63	12	4	1	0	0	1	0	0	0	0	0	0	0
10	003	016902	27	86	196	87	77	36	17	1	1	1	0	0	0	0	0	0	0	0	0
10	005	050101	38	235	465	213	203	72	22	10	5	0	1	3	0	0	0	0	0	0	0
10	005	050102	261	673	1019	563	357	218	81	27	18	7	4	1	3	1	0	0	0	0	0
10	005	050103	351	202	430	181	100	36	19	5	1	1	0	0	0	0	0	0	0	0	0
10	005	050200	109	205	307	182	167	74	36	13	5	2	1	1	0	1	0	0	0	0	0

10 005 050301	250	522	700	378	296	134	71	23	3	6	0	0	1	0	0	0	0	0	0
10 005 050302	84	207	454	277	222	104	52	11	5	2	0	0	0	0	0	0	0	0	0
10 005 050401	62	216	469	259	196	82	16	10	5	0	1	0	0	0	0	0	0	0	0
10 005 050402	211	964	1055	499	358	170	65	23	9	4	3	0	0	0	0	0	0	0	0
10 005 050403	93	249	341	216	157	81	30	7	6	4	2	2	0	0	0	0	0	0	0
10 005 050404	316	553	1005	606	472	209	84	18	10	6	1	1	0	0	0	0	0	0	0
10 005 050501	46	195	387	219	177	62	20	7	3	1	0	0	1	0	0	0	0	0	0
10 005 050502	230	569	712	395	364	191	74	53	35	17	17	8	3	0	2	0	1	1	0
10 005 050601	57	231	478	277	197	86	25	4	1	1	0	0	1	0	0	0	0	0	0
10 005 050602	179	516	586	238	165	76	22	11	4	1	0	0	0	1	0	0	0	0	0
10 005 050701	283	244	443	165	140	73	27	4	1	0	1	0	0	0	0	0	0	0	0
10 005 050702	3436	1049	2138	491	300	146	48	19	5	4	1	0	2	0	0	0	0	0	0
10 005 050801	127	302	364	196	138	49	18	8	4	0	0	0	0	0	0	0	0	0	0
10 005 050802	200	267	500	208	178	92	21	11	5	1	0	0	1	0	0	0	0	0	0
10 005 050803	861	439	871	278	227	102	18	11	0	2	0	1	0	0	0	0	0	0	0
10 005 050900	1581	815	1059	239	189	65	12	2	3	0	0	0	0	0	0	0	0	0	0
10 005 051001	1024	578	1113	298	236	72	28	7	1	1	0	0	0	0	0	0	0	0	0
10 005 051002	2386	679	1248	254	192	63	13	7	1	2	0	0	0	0	0	0	0	0	0
10 005 051003	794	406	565	182	135	62	22	8	2	0	1	0	0	0	0	0	0	0	0
10 005 051100	4529	690	674	112	57	24	7	2	0	1	1	0	0	0	0	0	0	0	0
10 005 051200	7647	619	1161	167	105	50	8	1	0	0	0	0	0	0	0	0	0	0	0
10 005 051301	885	362	896	204	161	62	19	4	0	0	0	1	1	0	0	0	0	0	0
10 005 051302	413	252	378	177	120	49	18	8	2	0	0	1	0	0	0	0	0	0	0
10 005 051303	782	338	588	174	111	41	13	2	0	1	0	0	0	0	0	0	0	0	0
10 005 051304	1533	443	972	195	131	33	17	6	1	1	0	0	0	0	0	0	0	0	0
10 005 051400	96	213	374	176	149	81	31	19	7	5	2	2	0	0	0	0	0	0	0
10 005 051500	180	344	546	302	213	100	41	17	3	2	3	3	0	0	1	0	0	0	0
10 005 051701	65	197	489	231	215	83	27	13	1	2	0	0	1	0	0	0	0	0	0
10 005 051702	138	347	672	347	289	119	33	13	5	2	0	0	0	0	0	0	0	0	0
10 005 051801	123	390	590	306	226	98	31	6	4	1	0	0	0	0	0	0	0	0	0
10 005 051802	174	349	535	239	221	105	40	22	3	2	1	0	0	0	0	0	0	0	0
10 005 051900	125	375	500	293	194	97	42	6	6	1	0	1	2	0	0	0	0	0	0
24 011 955000	104	201	312	173	175	96	39	22	8	4	1	3	0	1	0	0	0	0	0
24 011 955100	146	383	596	367	303	127	57	10	6	2	2	1	0	0	0	0	0	0	0
24 011 955200	134	390	640	365	326	154	47	15	5	2	1	0	0	0	0	0	0	0	0
24 011 955300	173	580	779	389	291	112	34	9	4	2	2	0	0	0	0	0	0	0	0
24 011 955400	64	113	291	134	148	46	12	4	2	0	0	0	0	0	0	0	0	0	0
24 011 955500	143	257	554	287	230	88	27	7	2	2	0	0	0	0	0	0	0	0	0
24 011 955600	135	464	643	319	266	103	41	11	7	2	1	0	0	0	0	0	0	0	0
24 015 030100	733	375	616	278	205	77	27	9	2	0	0	2	0	0	0	0	0	0	0
24 015 030200	389	391	738	357	315	119	35	15	5	2	2	0	0	1	0	0	0	0	0
24 019 970100	134	279	423	192	158	52	19	8	2	0	0	0	0	0	0	0	0	0	0
24 019 970200	178	418	601	340	299	136	51	13	5	0	0	2	0	0	0	0	0	0	0
24 019 970300	166	374	678	333	222	79	20	8	1	0	2	1	0	0	0	0	0	0	0
24 019 970400	204	635	528	227	142	51	10	12	9	0	0	1	0	0	0	0	0	0	0
24 019 970500	227	597	436	273	176	79	26	11	3	0	1	0	1	0	0	0	0	0	0
24 019 970600	166	446	505	297	171	62	17	7	0	0	0	0	0	0	0	0	0	0	0
24 019 970700	255	401	732	264	161	58	13	4	0	0	1	0	0	0	0	0	0	0	0
24 019 970800	343	224	394	124	82	40	6	5	0	1	1	0	0	0	0	0	0	0	0
24 019 970900	273	214	280	135	89	33	3	0	2	0	0	0	0	0	0	0	0	0	0
24 029 950100	232	293	467	199	170	72	24	9	3	2	0	1	0	0	0	0	0	0	0
24 029 950200	514	380	733	282	279	115	31	10	5	2	1	0	0	0	0	0	0	0	0
24 029 950300	280	815	752	255	157	50	13	6	1	1	0	0	0	0	0	0	0	0	0
24 029 950400	306	275	555	215	168	62	18	7	2	0	0	0	0	0	0	0	0	0	0
24 029 950500	375	365	509	172	134	41	12	2	1	0	0	0	0	0	0	0	0	0	0
24 035 810100	136	215	425	217	146	55	16	8	1	1	0	0	0	0	0	0	0	0	0
24 035 810200	70	187	304	163	149	71	20	5	1	1	0	0	0	0	0	0	0	0	0
24 035 810300	108	310	564	211	214	91	20	4	4	0	0	0	0	0	0	0	0	0	0
24 035 810400	110	305	564	249	257	89	38	3	3	1	0	0	0	0	0	0	0	0	0

24 035 810500	55	236	355	207	188	75	13	7	1	4	0	0	0	0	0	0	0	0	0
24 035 810600	204	268	794	283	270	134	40	7	3	0	0	0	0	0	0	0	0	0	0
24 035 810700	186	296	465	198	146	67	24	2	6	0	0	0	0	0	0	1	0	0	0
24 035 810800	166	382	619	365	310	144	38	14	2	4	0	0	1	0	0	0	0	0	0
24 035 810900	204	459	1097	485	494	188	52	17	10	2	1	0	0	0	0	0	0	0	0
24 035 811000	97	343	608	281	259	100	29	5	2	1	0	0	1	0	0	0	0	0	0
24 039 980101	262	618	480	287	203	86	35	10	5	4	1	0	0	0	0	0	0	0	0
24 039 980102	217	355	512	243	176	62	26	7	3	1	0	0	0	0	0	0	0	0	0
24 039 980200	396	241	372	124	90	41	17	2	1	2	0	0	0	0	0	0	0	0	0
24 039 980300	245	308	394	194	146	59	12	4	4	2	3	0	2	0	0	0	0	0	0
24 039 980400[‡]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24 039 980500	360	485	711	298	214	90	24	10	4	5	2	1	1	0	0	0	0	0	0
24 039 980600	236	453	463	228	152	55	20	5	5	1	2	0	0	0	0	0	0	0	0
24 041 960100	96	246	533	281	285	110	29	11	1	0	0	1	0	0	0	0	0	0	0
24 041 960200	222	354	700	166	137	48	12	7	0	0	0	0	0	0	0	0	0	0	0
24 041 960300	85	486	417	199	152	62	18	11	3	2	0	0	0	0	0	0	0	0	0
24 041 960400	222	1054	781	315	192	95	21	8	3	0	1	0	0	0	0	0	0	0	0
24 041 960500	143	505	840	417	338	134	38	11	3	1	1	0	0	0	0	0	0	0	0
24 041 960600	239	177	358	103	82	30	10	1	0	0	0	0	0	0	0	0	0	0	0
24 041 960700	456	540	838	210	150	58	19	1	4	0	0	0	0	0	0	0	0	0	0
24 041 960800	335	227	385	110	96	28	12	1	2	0	0	0	0	0	0	0	0	0	0
24 041 960900	340	393	800	289	223	96	26	7	1	1	0	0	0	0	0	0	0	0	0
24 045 000100	119	572	597	350	237	129	56	23	12	5	1	2	1	0	0	0	0	0	0
24 045 000200	70	280	269	120	79	33	11	9	1	0	2	0	0	0	0	0	0	0	0
24 045 000300	46	172	143	114	71	38	20	8	4	1	0	0	0	0	0	0	0	0	0
24 045 000400	119	652	516	278	213	82	18	7	3	3	0	2	0	0	0	0	0	0	0
24 045 000500	72	291	336	192	142	67	24	10	6	3	1	0	0	0	0	0	0	0	0
24 045 010101	81	617	761	415	285	107	37	18	6	3	1	0	0	0	0	0	0	0	0
24 045 010102	62	385	491	187	160	47	10	2	2	0	0	0	0	0	0	0	0	0	0
24 045 010200	138	470	563	386	300	142	69	15	15	0	1	2	0	0	0	0	0	0	0
24 045 010300	65	274	697	409	465	188	41	15	2	3	1	0	0	0	0	0	0	0	0
24 045 010400	114	654	737	324	297	119	21	7	3	2	4	0	0	0	0	0	0	0	0
24 045 010500	230	1054	1141	591	387	128	34	10	0	2	1	0	0	0	0	0	0	0	0
24 045 010602	159	516	1121	616	535	199	47	9	1	1	2	0	0	0	0	0	0	0	0
24 045 010603	93	379	736	378	328	122	36	20	5	1	1	0	0	0	0	0	0	0	0
24 045 010604	132	408	644	409	293	93	33	9	5	1	1	0	0	0	0	0	0	0	0
24 045 010701	145	289	450	230	209	80	27	6	1	3	1	0	0	0	0	0	0	0	0
24 045 010702	206	527	826	501	401	150	64	17	10	5	2	0	0	0	0	0	0	0	1
24 045 010800	320	461	816	418	313	108	27	7	5	0	0	0	0	0	0	0	0	0	0
24 047 990100	21756	1477	1638	358	197	78	23	7	1	0	0	1	0	0	1	0	0	0	0
24 047 990200	203	347	631	315	266	97	24	5	3	0	0	0	0	0	0	0	0	0	0
24 047 990300	113	459	455	289	213	96	29	6	7	0	0	2	1	0	0	0	0	0	0
24 047 990400	122	160	295	134	112	36	11	7	0	0	0	1	0	0	0	0	0	0	0
24 047 990500	1549	487	1350	301	221	91	22	5	2	1	0	1	1	0	0	0	0	0	0
24 047 990600	943	422	1138	320	211	72	20	2	2	0	0	0	0	0	0	0	0	0	0
24 047 990700	1298	481	843	264	219	75	21	4	4	0	1	0	0	0	0	0	0	0	0
24 047 990800	125	222	442	226	180	60	17	9	4	1	0	0	0	0	0	0	0	0	0
24 047 990900	115	304	349	179	108	57	18	7	4	1	1	0	1	0	0	0	0	0	0
24 047 991000	105	244	416	210	178	78	25	2	5	0	0	1	0	0	0	0	0	0	0
24 047 991100	200	577	592	338	284	129	33	18	5	1	4	1	0	0	0	0	0	0	0
51 001 990100	1875	700	832	301	164	58	8	3	1	1	0	0	0	0	0	0	0	0	0
51 001 990200	485	847	1065	604	510	207	65	34	13	12	2	2	1	1	0	0	0	0	0
51 001 990300	233	277	430	185	154	44	30	11	5	5	2	0	0	0	0	0	0	0	0
51 001 990400	339	501	623	367	296	143	71	25	11	8	3	2	2	1	0	0	0	0	0
51 001 990500	181	347	366	211	137	47	18	10	3	2	2	0	0	1	0	0	0	0	0
51 001 990600	356	619	697	343	218	81	18	6	3	0	1	0	0	0	0	0	0	0	0
51 001 990700	364	498	736	362	274	94	36	11	8	3	0	0	0	1	0	0	0	0	0

[‡]tract 980400 in Somerset County, MD contained no residential population in 2000 and was dropped from the analysis

51 001 990800	324	452	535	235	175	77	28	8	6	2	0	0	1	0	0	0	0	0	0
51 131 990100	359	476	636	292	197	72	24	14	1	1	0	1	0	0	0	0	0	0	0
51 131 990200	393	474	656	261	202	80	37	10	5	6	1	2	0	0	0	0	0	0	0
51 131 990300	444	615	594	277	205	108	45	15	8	4	2	0	0	0	0	0	0	0	0

Table 8: Number of housing units by housing unit size and tract

B Covariates used

state / county / tract	logist0	logist1	logist2	logist3	logistprop
10 001 040100	-1.273	-1.61907	-0.52961	-0.5736	0.269333
10 001 040201	-1.3029	-1.2712	-0.30418	-0.31393	0.633803
10 001 040202	-1.3029	-1.2712	-0.30418	-0.31393	0.633803
10 001 040203	-1.3029	-1.2712	-0.30418	-0.31393	0.633803
10 001 040400	-0.58766	-1.01543	-0.26688	-0.21706	0.71784
10 001 040500	-1.0423	-1.46505	-0.2243	-0.12024	0.54932
10 001 040600	-1.02029	-0.42973	-0.08224	-0.21131	0.904456
10 001 040700	-1.08219	-1.54464	-0.67188	-0.4873	0.693147
10 001 040800	-1.42742	-0.92716	-0.33386	-0.13119	0.332382
10 001 040900	-1.23245	-0.09421	0.300486	0.144581	0.548566
10 001 041000	-1.20798	-1.06352	-0.17605	-0.22253	0.572631
10 001 041100	-1.95148	-3.64087	-1.62924	-0.58901	0.972732
10 001 041200	-1.2417	-1.23752	-0.46557	-0.32886	0.411855
10 001 041300	-1.13423	-0.5066	0.066797	-0.10536	0.402345
10 001 041400	-1.1333	-0.79116	0.028581	-0.33881	0.359084
10 001 041500	-1.80163	-1.70397	-0.40352	-0.32182	0.695387
10 001 041600	-1.7262	-1.6484	-0.11872	-0.43457	0.555526
10 001 041701	-1.48053	-1.49428	-0.44594	-0.31626	0.415081
10 001 041702	-1.48053	-1.49428	-0.44594	-0.31626	0.415081
10 001 041801	-1.4489	-1.30915	-0.35914	-0.38038	0.121782
10 001 041802	-1.4489	-1.30915	-0.35914	-0.38038	0.121782
10 001 041900	-1.4624	-1.61187	-0.49485	-0.46788	0.314216
10 001 042000	-1.23736	-1.65823	-0.63173	-0.5221	0.202712
10 001 042100	-1.31251	-1.41923	-0.31356	-0.28986	0.30305
10 001 042201	-1.16303	-1.67019	-0.46472	-0.34935	0.510467
10 001 042202	-1.16303	-1.6702	-0.46473	-0.34936	0.510476
10 001 042400	-0.5724	-1.35473	-0.11675	0.00551	0.623786
10 001 042500	-1.11973	-0.64064	-0.23832	-0.36546	0.228259
10 001 042600	-1.42916	-1.49789	-0.02899	-0.17185	0.731466
10 001 042700	-1.11011	-1.0033	-0.40547	-0.20271	0.272507
10 001 042800	-1.26804	-1.40127	-0.51045	-0.23436	0.64755
10 001 042900	-1.20119	-1.56255	-0.59289	-0.22642	0.381251
10 001 043000	-1.19658	-1.15568	-0.28172	-0.21035	0.434342
10 001 043100	-1.13121	-1.62892	-0.58179	-0.16579	0.81093
10 003 016601	-1.30884	-1.46553	-0.64543	-0.39453	0.652117
10 003 016602	-1.21631	-2.03708	-0.73216	-0.47987	0.723878
10 003 016603	-1.21534	-2.04534	-0.7331	-0.48077	0.724623
10 003 016604	-1.34914	-1.29737	-0.60958	-0.35759	0.620693
10 003 016801	-1.21347	-1.69867	-0.4634	-0.47776	0.542067
10 003 016802	-1.21347	-1.69867	-0.4634	-0.47776	0.542067
10 003 016901	-1.32424	-1.66627	-0.41468	-0.43614	0.587183
10 003 016902	-1.32424	-1.66627	-0.41468	-0.43614	0.587183
10 005 050101	-0.79932	-1.38806	-0.25804	-0.32736	0.47324

10 005 050102	-0.79932	-1.38806	-0.25804	-0.32736	0.47324
10 005 050103	-0.79932	-1.38806	-0.25804	-0.32736	0.47324
10 005 050200	-0.98346	-1.35239	-0.7126	-0.44848	0.349184
10 005 050301	-1.00987	-1.31019	-0.32105	-0.32051	0.328752
10 005 050302	-1.00987	-1.31019	-0.32105	-0.32051	0.328752
10 005 050401	-1.12208	-1.34703	-0.30938	-0.26993	0.511749
10 005 050402	-1.12208	-1.34703	-0.30938	-0.26993	0.511749
10 005 050403	-1.12195	-1.34695	-0.30924	-0.26987	0.511624
10 005 050404	-1.12208	-1.34703	-0.30938	-0.26993	0.511749
10 005 050501	-1.01589	-1.21337	-0.28241	-0.23291	0.57089
10 005 050502	-1.01589	-1.21337	-0.28241	-0.23291	0.57089
10 005 050601	-0.8549	-1.10636	-0.1882	-0.1283	0.527979
10 005 050602	-0.8549	-1.10636	-0.1882	-0.1283	0.527979
10 005 050701	0.199329	-1.29536	0.454844	0.015748	0.501796
10 005 050702	0.199562	-1.29234	0.454722	0.014693	0.502513
10 005 050801	-0.51931	-1.26919	-0.19699	-0.28676	0.532292
10 005 050802	-0.51931	-1.26919	-0.19699	-0.28676	0.532292
10 005 050803	-0.51931	-1.26919	-0.19699	-0.28676	0.532292
10 005 050900	-0.0572	-0.89128	0.262942	0.022557	0.6449
10 005 051001	0.121618	-1.02645	0.379056	-0.16913	0.607718
10 005 051002	0.121618	-1.02645	0.379056	-0.16913	0.607718
10 005 051003	0.121618	-1.02645	0.379056	-0.16913	0.607718
10 005 051100	0.593812	-0.34324	0.892829	0.126477	0.596819
10 005 051200	0.963078	-1.01555	0.966441	0.229574	0.661398
10 005 051301	0.079594	-1.30211	0.283923	0.030021	0.795338
10 005 051302	0.079594	-1.30211	0.283923	0.030021	0.795338
10 005 051303	0.08412	-1.30162	0.284847	0.030209	0.795216
10 005 051304	0.079698	-1.3021	0.283944	0.030025	0.795335
10 005 051400	-0.89787	-1.27841	-0.26311	-0.20972	0.119474
10 005 051500	-0.88757	-1.60421	-0.44056	-0.21927	0.289411
10 005 051701	-1.10171	-1.58329	-0.33693	-0.21191	0.750185
10 005 051702	-1.05428	-1.59655	-0.40364	-0.15169	0.796032
10 005 051801	-1.03678	-1.28591	-0.20215	-0.22553	0.418265
10 005 051802	-1.03678	-1.28591	-0.20215	-0.22553	0.418265
10 005 051900	-1.07918	-1.43423	-0.33501	-0.28662	0.325668
24 011 955000	-1.1821	-1.46674	-0.4317	-0.48752	0.310447
24 011 955100	-1.17412	-1.29377	-0.29881	-0.11846	0.337017
24 011 955200	-1.15104	-1.42583	-0.44764	-0.32559	0.461723
24 011 955300	-1.16505	-1.05842	-0.14537	-0.31312	0.586972
24 011 955400	-1.04402	-1.71816	-0.5042	-0.35114	0.430036
24 011 955500	-0.98674	-1.43547	-0.33713	-0.26469	0.562527
24 011 955600	-1.16425	-1.10553	-0.21074	-0.27071	0.411596
24 015 030100	-0.24172	-1.30218	0.013501	-0.25378	0.588477
24 015 030200	-0.57673	-1.53212	-0.21319	-0.38329	0.607257
24 019 970100	-0.84433	-1.14419	-0.26654	-0.28611	0.560713
24 019 970200	-0.95764	-1.37982	-0.52818	-0.34134	0.405465
24 019 970300	-0.91673	-1.44533	-0.12117	-0.03604	0.763405
24 019 970400	-0.9398	-0.58144	0.138402	-0.06252	0.763765
24 019 970500	-0.98756	-0.67056	-0.38792	-0.26706	0.470004
24 019 970600	-1.06482	-1.09538	-0.11263	0.046596	0.609009
24 019 970700	-0.73444	-1.28864	0.188596	0.024103	0.96953
24 019 970800	-0.30973	-1.22118	0.134781	-0.03433	1.12592
24 019 970900	-0.35539	-1.08163	-0.02055	0.101783	0.814508
24 029 950100	-0.65282	-1.31053	-0.33647	-0.41871	0.283306
24 029 950200	-0.5838	-1.36451	-0.24769	-0.35378	0.51521
24 029 950300	-0.88461	-0.61149	0.126294	-0.09409	0.610909
24 029 950400	-0.70426	-1.43527	0.009372	-0.29208	0.660782
24 029 950500	-0.58978	-1.01371	0.02076	-0.03037	0.859697
24 035 810100	-0.74819	-1.38236	-0.35317	-0.16593	0.324152

24 035 810200	-1.0196	-1.69105	-0.53178	-0.33489	0.451985
24 035 810300	-1.08052	-1.37671	-0.15833	-0.30228	0.825792
24 035 810400	-1.15585	-1.15467	-0.2265	-0.04565	0.458002
24 035 810500	-1.15507	-1.37932	-0.26204	-0.16248	0.449954
24 035 810600	-0.91524	-1.81573	0.001528	-0.27076	0.456617
24 035 810700	-0.67071	-1.4433	-0.20654	-0.24039	0.355455
24 035 810800	-0.74768	-1.59302	-0.36817	-0.33922	0.659474
24 035 810900	-0.989	-1.85982	-0.30675	-0.27549	0.885333
24 035 811000	-1.07636	-1.40264	-0.2646	-0.38692	0.806262
24 039 980101	-0.92176	-1.00719	-0.18808	-0.12584	0.402092
24 039 980102	-0.92176	-1.00719	-0.18808	-0.12584	0.402092
24 039 980200	-0.40193	-1.16904	-0.12579	-0.10131	0.634896
24 039 980300	-0.77383	-1.21227	-0.13954	-0.32513	0.309464
24 039 980500	-0.72721	-1.21367	0.03473	-0.01047	0.486811
24 039 980600	-0.8957	-0.76092	-0.27647	-0.11058	0.394282
24 041 960100	-1.10443	-1.7618	-0.39132	-0.35449	0.555161
24 041 960200	-0.68649	-1.4144	0.244278	-0.40344	0.855428
24 041 960300	-1.14578	-0.59694	0.064061	0.158089	0.573524
24 041 960400	-1.15405	-0.59717	0.153738	0.05782	0.786673
24 041 960500	-1.14856	-1.43813	-0.12334	-0.08088	0.752519
24 041 960600	-0.47108	-1.35001	0.406858	-0.07538	0.747505
24 041 960700	-0.62571	-1.04744	0.512099	0.089231	0.841389
24 041 960800	-0.44731	-1.10621	0.129887	-0.18544	0.814265
24 041 960900	-0.69818	-1.16093	0.133744	-0.11255	0.655168
24 045 000100	-1.25746	-0.94995	-0.35634	-0.2863	0.568472
24 045 000200	-0.91867	-0.56761	-0.06372	-0.18987	0.586284
24 045 000300	-1.17386	-0.89355	-0.48893	-0.31508	0.099372
24 045 000400	-1.26603	-0.64267	-0.04725	-0.07656	0.581891
24 045 000500	-1.31903	-1.11343	-0.10303	-0.1354	0.535882
24 045 010101	-1.35373	-1.48416	-0.11953	0.050644	0.675129
24 045 010102	-1.22416	-1.03936	0.0743	0.209051	1.070193
24 045 010200	-1.24853	-1.40738	-0.61996	-0.31707	0.13644
24 045 010300	-1.03641	-1.37988	-0.43258	-0.54598	0.831197
24 045 010400	-1.1261	-1.223	-0.05329	-0.23397	0.543867
24 045 010500	-1.20955	-0.8958	-0.02077	-0.06862	0.713955
24 045 010602	-1.16789	-1.67225	-0.47	-0.30291	0.981719
24 045 010603	-1.05463	-1.45621	-0.33033	-0.12212	0.629691
24 045 010604	-1.05463	-1.45621	-0.33033	-0.12212	0.629691
24 045 010701	-0.92256	-1.24009	-0.43233	-0.23879	0.42028
24 045 010702	-1.00886	-1.31824	-0.39427	-0.20435	0.486481
24 045 010800	-0.63283	-1.18941	-0.20802	-0.07059	0.68784
24 047 990100	0.92682	-0.57972	0.734467	0.29502	0.692864
24 047 990200	-0.74006	-1.43028	-0.22469	-0.01858	0.529402
24 047 990300	-0.88137	-0.99446	-0.33008	-0.27339	0.423907
24 047 990400	-0.65612	-1.3138	-0.09029	-0.28978	0.29822
24 047 990500	0.160148	-1.77843	0.357218	-0.02174	0.571786
24 047 990600	-0.00673	-1.62123	0.293547	-0.16508	0.884202
24 047 990700	-0.04085	-1.13273	0.190129	-0.11575	0.990268
24 047 990800	-0.90304	-1.44675	-0.3474	-0.3119	0.516973
24 047 990900	-1.1706	-1.096	-0.23078	0.070287	0.415221
24 047 991000	-0.88484	-1.41694	-0.30002	-0.35753	0.602175
24 047 991100	-1.05219	-0.92191	-0.21936	-0.32609	0.54024
51 001 990100	-0.06394	-0.7582	0.436481	0.296662	1.04835
51 001 990200	-0.74785	-1.08177	-0.29759	-0.26145	0.285405
51 001 990300	-0.74348	-0.98165	-0.01966	-0.33898	0.192372
51 001 990400	-0.96213	-1.14051	-0.34166	-0.22259	0.23935
51 001 990500	-0.90887	-0.93922	-0.16705	-0.14505	0.464114
51 001 990600	-0.84581	-0.87609	0.021673	-0.10338	0.669617
51 001 990700	-0.82088	-1.02945	-0.19924	-0.31869	0.439481

51 001 990800	-0.81218	-1.00744	-0.17655	-0.26387	0.324496
51 131 990100	-0.65931	-1.08766	-0.04912	-0.18232	0.257829
51 131 990200	-0.65003	-0.99857	-0.11463	-0.33079	0.239502
51 131 990300	-0.811	-0.81921	-0.16227	-0.39623	0.205322

Table 9: Covariates by tract

C Tracts and Their Neighbors

state / county / tract	ID	#	Neighbor ID's
10 001 0401	1	7	15 , 51 , 57 , 112 , 113 , 117 , 127
10 001 040201	112	5	1 , 113 , 114 , 127 , 128
10 001 040202	113	5	1 , 3 , 112 , 114 , 117
10 001 040203	114	6	2 , 3 , 112 , 113 , 127 , 128
10 001 0404	2	7	3 , 8 , 9 , 18 , 114 , 120 , 128
10 001 0405	3	7	2 , 4 , 7 , 8 , 113 , 114 , 117
10 001 0406	4	5	3 , 5 , 6 , 7 , 117
10 001 0407	5	4	4 , 6 , 12 , 117
10 001 0408	6	5	4 , 5 , 7 , 11 , 12
10 001 0409	7	6	3 , 4 , 6 , 8 , 11 , 12
10 001 0410	8	6	2 , 3 , 7 , 9 , 10 , 11
10 001 0411	9	6	2 , 8 , 10 , 18 , 119 , 120
10 001 0412	10	6	8 , 9 , 11 , 14 , 116 , 119
10 001 0413	11	7	6 , 7 , 8 , 10 , 12 , 13 , 14
10 001 0414	12	8	5 , 6 , 7 , 11 , 13 , 115 , 117 , 118
10 001 0415	13	4	11 , 12 , 14 , 115
10 001 0416	14	5	10 , 11 , 13 , 115 , 116
10 001 041701	115	7	12 , 13 , 14 , 17 , 116 , 118 , 119
10 001 041702	116	5	10 , 14 , 17 , 115 , 119
10 001 041801	117	8	1 , 3 , 4 , 5 , 12 , 15 , 113 , 118
10 001 041802	118	6	12 , 15 , 16 , 17 , 115 , 117
10 001 0419	15	6	1 , 16 , 33 , 57 , 117 , 118
10 001 0420	16	6	15 , 17 , 22 , 33 , 34 , 118
10 001 0421	17	7	16 , 22 , 115 , 116 , 118 , 119 , 120
10 001 042201	119	6	9 , 10 , 17 , 115 , 116 , 120
10 001 042202	120	7	2 , 9 , 17 , 18 , 21 , 22 , 119
10 001 0424	18	7	2 , 9 , 19 , 20 , 21 , 120 , 131
10 001 0425	19	5	18 , 20 , 129 , 130 , 131
10 001 0426	20	6	18 , 19 , 21 , 23 , 129 , 130
10 001 0427	21	5	18 , 20 , 22 , 23 , 120
10 001 0428	22	8	16 , 17 , 21 , 23 , 24 , 25 , 34 , 120
10 001 0429	23	8	20 , 21 , 22 , 24 , 25 , 129 , 132 , 133
10 001 0430	24	3	22 , 23 , 25
10 001 0431	25	7	22 , 23 , 24 , 34 , 36 , 132 , 133
10 003 016601	121	6	40 , 41 , 122 , 124 , 125 , 126
10 003 016602	122	3	121 , 123 , 124
10 003 016603	123	3	122 , 124 , 126
10 003 016604	124	5	121 , 122 , 123 , 125 , 126
10 003 016801	125	7	40 , 51 , 121 , 124 , 126 , 127 , 128
10 003 016802	126	6	121 , 123 , 124 , 125 , 127 , 128
10 003 016901	127	6	1 , 112 , 114 , 125 , 126 , 128
10 003 016902	128	6	2 , 112 , 114 , 125 , 126 , 127
10 005 050101	129	6	19 , 20 , 23 , 26 , 130 , 133

10 005 050102	130	5	19 , 20 , 26 , 129 , 131
10 005 050103	131	6	18 , 19 , 26 , 27 , 130 , 144
10 005 0502	26	8	129 , 130 , 131 , 133 , 138 , 139 , 144 , 145
10 005 050301	132	9	23 , 25 , 36 , 37 , 39 , 133 , 134 , 135 , 137
10 005 050302	133	7	23 , 25 , 26 , 129 , 132 , 137 , 138
10 005 050401	134	7	39 , 42 , 132 , 135 , 136 , 156 , 165
10 005 050402	135	4	132 , 134 , 136 , 137
10 005 050403	136	5	134 , 135 , 137 , 154 , 156
10 005 050404	137	8	132 , 133 , 135 , 136 , 138 , 140 , 154 , 156
10 005 050501	138	5	26 , 133 , 137 , 139 , 140
10 005 050502	139	5	26 , 138 , 140 , 142 , 145
10 005 050601	140	9	31 , 137 , 138 , 139 , 141 , 142 , 143 , 154 , 155
10 005 050602	141	3	31 , 140 , 143
10 005 050701	142	5	139 , 140 , 143 , 145 , 147
10 005 050702	143	9	28 , 29 , 31 , 140 , 141 , 142 , 147 , 148 , 150
10 005 050801	144	5	26 , 27 , 131 , 145 , 146
10 005 050802	145	7	26 , 27 , 139 , 142 , 144 , 146 , 147
10 005 050803	146	4	27 , 144 , 145 , 147
10 005 0509	27	7	28 , 131 , 144 , 145 , 146 , 147 , 149
10 005 051001	147	7	27 , 142 , 143 , 145 , 146 , 148 , 149
10 005 051002	148	4	28 , 143 , 147 , 149
10 005 051003	149	4	27 , 28 , 147 , 148
10 005 0511	28	6	27 , 29 , 143 , 148 , 149 , 150
10 005 0512	29	6	28 , 90 , 143 , 150 , 152 , 153
10 005 051301	150	6	28 , 29 , 31 , 143 , 151 , 152
10 005 051302	151	5	30 , 31 , 150 , 152 , 153
10 005 051303	152	4	29 , 150 , 151 , 153
10 005 051304	153	6	29 , 30 , 90 , 91 , 151 , 152
10 005 0514	30	5	31 , 91 , 151 , 153 , 155
10 005 0515	31	7	30 , 140 , 141 , 143 , 150 , 151 , 155
10 005 051701	154	6	136 , 137 , 140 , 155 , 156 , 157
10 005 051702	155	9	30 , 31 , 32 , 91 , 140 , 154 , 157 , 163 , 164
10 005 051801	156	8	32 , 42 , 134 , 136 , 137 , 154 , 157 , 165
10 005 051802	157	4	32 , 154 , 155 , 156
10 005 0519	32	5	155 , 156 , 157 , 165 , 166
24 011 9550	33	5	15 , 16 , 34 , 57 , 60
24 011 9551	34	8	16 , 22 , 25 , 33 , 35 , 36 , 57 , 60
24 011 9552	35	5	34 , 36 , 37 , 60 , 71
24 011 9553	36	5	25 , 34 , 35 , 37 , 132
24 011 9554	37	6	35 , 36 , 38 , 39 , 71 , 132
24 011 9555	38	6	37 , 39 , 43 , 71 , 75 , 79
24 011 9556	39	6	37 , 38 , 42 , 43 , 132 , 134
24 015 0301	40	5	41 , 51 , 52 , 121 , 125
24 015 0302	41	2	40 , 121
24 019 9701	42	7	39 , 43 , 44 , 50 , 134 , 156 , 165
24 019 9702	43	5	38 , 39 , 42 , 44 , 79
24 019 9703	44	5	42 , 43 , 48 , 50 , 79
24 019 9704	45	4	46 , 47 , 48 , 79
24 019 9705	46	3	45 , 47 , 48
24 019 9706	47	3	45 , 46 , 48
24 019 9707	48	7	44 , 45 , 46 , 47 , 49 , 50 , 79
24 019 9708	49	2	48 , 50
24 019 9709	50	7	42 , 44 , 48 , 49 , 66 , 89 , 165
24 029 9501	51	6	1 , 40 , 52 , 56 , 57 , 125
24 029 9502	52	5	40 , 51 , 53 , 54 , 56
24 029 9503	53	4	52 , 54 , 56 , 58
24 029 9504	54	5	52 , 53 , 55 , 58 , 59
24 029 9505	55	1	54
24 035 8101	56	5	51 , 52 , 53 , 57 , 58

24 035 8102	57	8	1 , 15 , 33 , 34 , 51 , 56 , 58 , 60
24 035 8103	58	6	53 , 54 , 56 , 57 , 59 , 60
24 035 8104	59	4	54 , 58 , 60 , 61
24 035 8105	60	8	33 , 34 , 35 , 57 , 58 , 59 , 61 , 71
24 035 8106	61	5	59 , 60 , 62 , 71 , 72
24 035 8107	62	3	61 , 63 , 65
24 035 8108	63	3	62 , 64 , 65
24 035 8109	64	2	63 , 65
24 035 8110	65	3	62 , 63 , 64
24 039 980101	158	5	67 , 87 , 99 , 159 , 162
24 039 980102	159	6	66 , 67 , 87 , 89 , 158 , 162
24 039 9802	66	4	50 , 67 , 106 , 159
24 039 9803	67	7	66 , 68 , 69 , 99 , 100 , 158 , 159
24 039 9804	68	1	67
24 039 9805	69	5	67 , 70 , 100 , 102 , 103
24 039 9806	70	1	69
24 041 9601	71	7	35 , 37 , 38 , 60 , 61 , 72 , 75
24 041 9602	72	8	61 , 71 , 73 , 74 , 75 , 76 , 77 , 79
24 041 9603	73	3	72 , 74 , 75
24 041 9604	74	3	72 , 73 , 75
24 041 9605	75	6	38 , 71 , 72 , 73 , 74 , 79
24 041 9606	76	3	72 , 77 , 79
24 041 9607	77	3	72 , 76 , 78
24 041 9608	78	1	77
24 041 9609	79	8	38 , 43 , 44 , 45 , 48 , 72 , 75 , 76
24 045 0001	80	6	81 , 83 , 84 , 88 , 160 , 161
24 045 0002	81	6	80 , 82 , 83 , 84 , 85 , 160
24 045 0003	82	4	81 , 83 , 85 , 86
24 045 0004	83	7	80 , 81 , 82 , 84 , 86 , 87 , 88
24 045 0005	84	6	80 , 81 , 83 , 87 , 88 , 161
24 045 010101	160	7	80 , 81 , 85 , 161 , 162 , 163 , 166
24 045 010102	161	6	80 , 84 , 88 , 160 , 162 , 163
24 045 0102	85	5	81 , 82 , 86 , 160 , 166
24 045 0103	86	6	82 , 83 , 85 , 87 , 89 , 166
24 045 0104	87	8	83 , 84 , 86 , 88 , 89 , 158 , 159 , 162
24 045 0105	88	6	80 , 83 , 84 , 87 , 161 , 162
24 045 010602	162	11	87 , 88 , 91 , 97 , 99 , 158 , 159 , 160 , 161 , 163 , 164
24 045 010603	163	6	155 , 160 , 161 , 162 , 164 , 166
24 045 010604	164	4	91 , 155 , 162 , 163
24 045 010701	165	7	32 , 42 , 50 , 89 , 134 , 156 , 166
24 045 010702	166	7	32 , 85 , 86 , 89 , 160 , 163 , 165
24 045 0108	89	6	50 , 86 , 87 , 159 , 165 , 166
24 047 9901	90	5	29 , 91 , 95 , 96 , 153
24 047 9902	91	12	30 , 90 , 92 , 93 , 94 , 95 , 96 , 97 , 153 , 155 , 162 , 164
24 047 9903	92	2	91 , 93
24 047 9904	93	6	91 , 92 , 94 , 95 , 96 , 97
24 047 9905	94	3	91 , 93 , 95
24 047 9906	95	5	90 , 91 , 93 , 94 , 96
24 047 9907	96	8	90 , 91 , 93 , 95 , 97 , 99 , 101 , 102
24 047 9908	97	6	91 , 93 , 96 , 98 , 99 , 162
24 047 9909	98	2	97 , 99
24 047 9910	99	8	67 , 96 , 97 , 98 , 100 , 102 , 158 , 162
24 047 9911	100	4	67 , 69 , 99 , 102
51 001 9901	101	2	96 , 102
51 001 9902	102	7	69 , 96 , 99 , 100 , 101 , 103 , 104
51 001 9903	103	4	69 , 102 , 104 , 105
51 001 9904	104	5	102 , 103 , 105 , 107 , 108
51 001 9905	105	4	103 , 104 , 106 , 107
51 001 9906	106	3	66 , 105 , 107

51 001 9907	107	5	104 , 105 , 106 , 108 , 109
51 001 9908	108	3	104 , 107 , 109
51 131 9901	109	4	107 , 108 , 110 , 111
51 131 9902	110	2	109 , 111
51 131 9903	111	2	109 , 110

Table 10: Tract neighbors by ID