Housing Affordability for Low Income Households in New Jersey, Maryland and West Virginia: a Comparative Analysis

George Obsekov
American College of Radiology Research Center, Philadelphia PA
Claudia Hauck
The Rutgers State University of New Jersey, New Brunswick NJ

I. Introduction

This project is designed to analyze the odds of housing affordability for low income families in New Jersey (NJ), Maryland (MD) and West Virginia (WV). The purpose of the analysis is to compare housing and rental statistics for NJ against two other states, in order to establish which state is most affordable for low-income households and what variables are important in predicting the odds of affordability. We define a household as low-income according to how the household income compares to the median income of a 4-person household.

The definition of low income household and federal poverty thresholds were created from information outside of the Public Use Micro Data Set and presented in Table 1 and 2.

Persons in household	Threshold for low-income (as % of median of a four-person household)
2	80%
3	90%
4	80%
5	108%
6	116%

Table 1: Definition of Low Income according to the family median income

Persons in household	Poverty threshold in dollars (defines POVG variable)
1	$ 8,350
2	11,250
3	14,150
4	17,050
5	19,950
6	22,850
7	25,750
8	28,650
9 or more	2900*(persons -8) + 28650 To compute the threshold for households of more than 8 members, for each additional person above 8.add $2900 to the threshold for 8 members.

Table 2: Federal poverty thresholds, based on number of persons in household

The state median incomes for family of 4 people in 2000 are: $65,500 (NJ), $62,800 (MD) and 37,450 (WV).

In order to compare affordability between states, we constructed six subsets with 387 to 2193 observations on housing units, by sub-setting the observations by state and by owner-occupied/renter-occupied status of the property, and choosing only new properties to analyze. Several predictor variables of housing affordability were considered; these variables are described in Table 3.

Variables considered in logistic regression	Name of variable in data set	Definitions
Response variable: Affordability	Aff Aff_enet	If percentage of income (GRAPI or SMOCAPI‡) going to housing costs is < 30%, then Aff = Aff_enet = 1 (the unit is affordable for low-income households). Otherwise, it is unaffordable, and Aff = 0, while Aff_enet = -1.
Persons living in residence	PERSONS	
Building size	BLDGSZ	1= mobile home, 2=one-family detached house, 3=one-family attached house, 4-and-higher = apartment in buildings of increasing size. Also created two binary variables representing mobile homes and apartment buildings.
Year the building was built	YRBUILT	YRBUILT / Years built 1 / 1999-2000 2 / 1995-1998 3 / 1990-1994 4 / 1980-1989 5 / 1970-1979
Number of rooms	ROOMS	Number of rooms in the household.
Number of vehicles	VEHICL	Number of Vehicles.
Gender	SEX	1=Male, 2=Female.
Age	AGE	Age of the head of the household
Senior citizen (binary)*	senior	If AGE >= 62, then senior = 1.
Education level	EDUC	Coded 1 through
Graduated from college/high school (binary)*	educoll/eduhs	If EDUC >= 13(9), then the person graduated from college (high school) and educoll /eduhs= 1.
Transformation of household income	lhinc	Transformation of household income
Travel time to work for head of household	TRVTIME	Measurement is in minutes
Whether travel time to work is good (binary)*	trvtime_good	If TRVTIME <= (the median TRVTIME for the state and ownership status), then travel time is deemed as "good," and trvtime_good = 1.
Household income is below the year 2000 federal poverty levels (binary)*	povg	If FINC <= poverty level for the number of persons in the household, then the household is deemed as being impoverished, as well as low-income. See Table 2 for poverty thresholds.
Single mother as head-of-household (binary)*	smother	If the head-of-household code (HHT) is equal to 3, then this is a household with a female at the head, and more than one person in the household. This is interpreted as the person being a single mother, and smother = 1 for this record.
Crowding condition in household (binary)	uncrowd	If the number of persons is more than the number of rooms, then the household is crowded and uncrowd = 1.
Selected Monthly Owner Costs as a Percentage of Household Income	SMOCAPI	Percentage of income going to housing cost.

Table 3: Variables used in the prediction of affordability of low-income housing

Our response variable was affordability (1=affordable, 0=unaffordable) based on the percentage of income going towards mortgage or rent. If this percentage is no more than 30%, the low-income housing unit was categorized as affordable.

As the first step, we compared income distributions of the states in pairs using parametric plots of the percentile income with the income value as a link. We also compared the distribution of property values for owner-occupied houses and the distribution of rents for rental units for each state using similar percentile plots. Next, we compared basic state statistics such as median income, proportion of low income households in the state, and various proportions that indicate quality and affordability of units.

After descriptive statistics, we constructed regression models to predict the odds of affordability. We then used LASSO and Elastic net to choose the most significant predictors from these models, and referred to the Cp-Mallow's statistic to help us choose

smaller subset. In order to select a cutoff lambda parameter for Elastic net, we analyzed cross-validation error plot from LASSO, and re-ran LASSO with this lambda to determine whether our choice of variables should change. This methodology led us to our definitions of the reduced sets of predictor variables. To confirm our new models, we examined the cross-validation error from Lasso with the final set of variables to be sure that it reached its minimum quickly near the lambda of 1. Finally, we corroborated the variable selection results with output from stepwise regression.

II. Results

1. Income, Rent and Household Value Distribution Analysis

In general, NJ has the highest income of the three states in all percentiles, followed by MD and WV. The gap between the percentiles of NJ and MD gets wider as income increases. For owners the difference between income distributions for WV and other two states is wider than it is for renters.

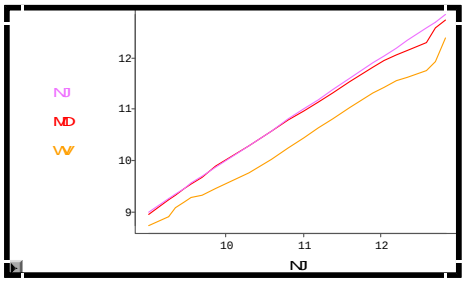


Figure 1: Distribution by income for NJ vs. MD & WV for all residents

In general, NJ has the highest rent distribution of three states in all percentiles, with MD coming in as a close second; WV has a much lower distribution. The gap between percentile of NJ and MD gets wider as rent value increases. For WV this is also true, but to a much greater extent.

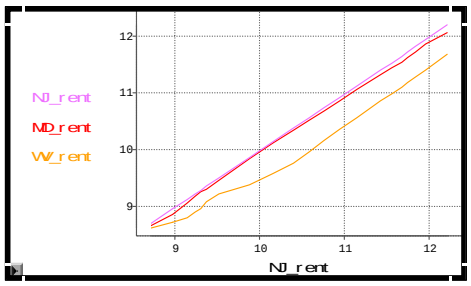


Figure 2: Distribution by income for NJ vs. MD & WV for rentals only

On the other hand, the trend of the percentile gaps is the opposite of that for household values: as household values increase, the gap decreases between each state and NJ.

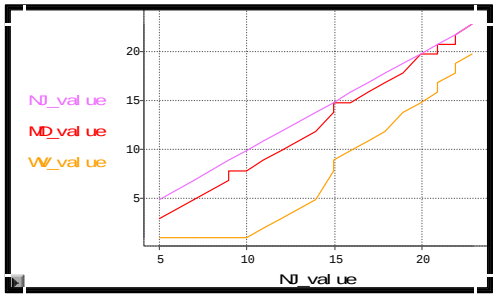


Figure 3: Distribution by household value for NJ vs. MD & WV

2. Percentage of income going toward housing costs

Table 4 demonstrates the percentages of buildings built in the years 1970 through 2000, which are affordable.

	Year built*				
	1999-2000	1995-1998	1990-1994	1980-1989	1970-1979
Maryland	72.42	75.60	79.02	80.81	80.59
New Jersey	73.32	73.28	74.07	75.23	76.06
West Virginia	75.83	77.97	81.08	85.38	85.62

Table 4: Percentage of owner-occupied housing that is affordable

Affordability of housing can also be defined more precisely by the percentage of income going toward housing costs. Following graphs display the distribution of these proportions for the owner-occupied units included in this report, for low-income owners.

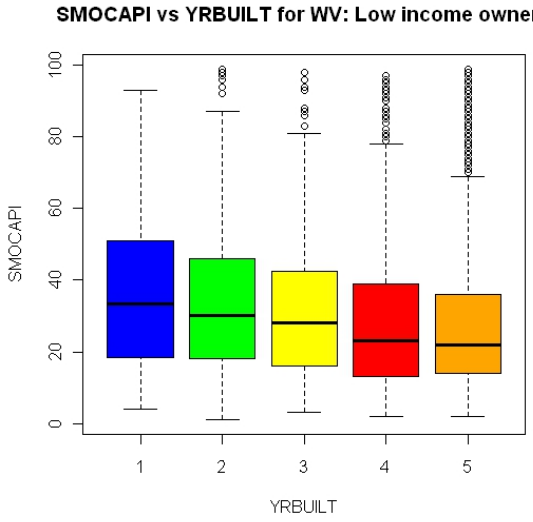


Figure 4: Box plot of SMOCAPI vs. YRBUILT for WV (Low income owners)

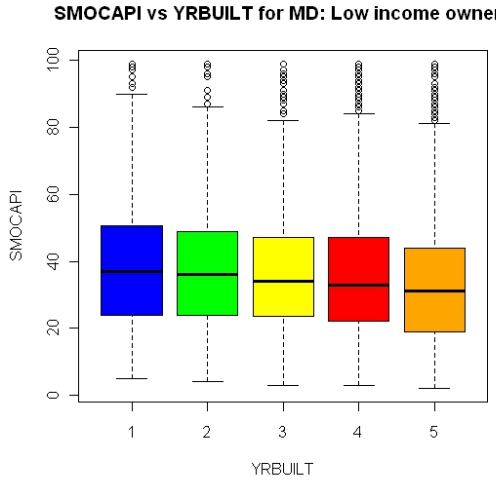


Figure 5: Box plot of SMOCAPI vs. YRBUILT for MD (Low income owners)

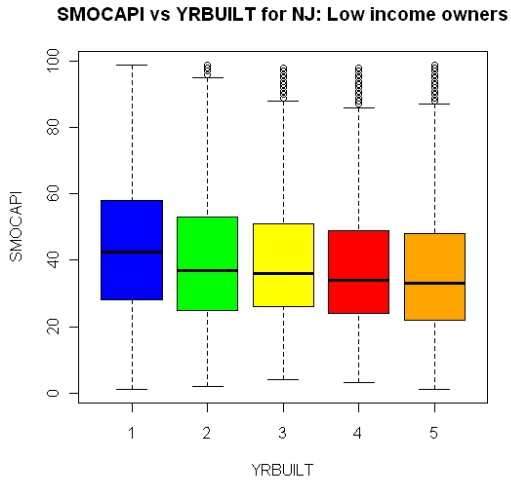


Figure 6: Box plot of SMOCAPI vs. YRBUILT for NJ (Low income owners)

As observed in these tables and figures, WV has consistently had the highest percentage of affordable housing for both the overall group of owners and group of owners living in low-income housing. NJ has consistently had the lowest percentage. Even as the percentage of affordable housing decreased for all three states since 1970, this comparison remains true.

3. Comparisons of state-wise proportions

As the percentile plots indicated, the median income for NJ is higher than for MD and WV. The proportion of low-income households is the highest in WV and lowest in MD. The proportion of units that are new, low-income, affordable and not crowded is the lowest for NJ and the highest for WV. We also observed that the proportion of new units

that are not crowded, are affordable and are occupied by low income families is the highest for WV and the lowest for NJ. The percentage of senior citizens living in NJ who are in low-income households and live in unaffordable housing are higher than for other states. WV has the highest percentage of owners living in mobile homes and a lower percentage living in one-family houses; it has the highest percentage of low-income renters living in mobile homes and one-family detached homes. Additionally, WV also has the highest percentage of both owner- and renter-occupied low-income housing that is affordable. NJ has the lowest percentage.

STATE	Percent of new units, occupied by low income families and are affordable	Percent of units that are new, low income, affordable and un-crowded	Percent that are low-income	Percent of low income, that are close to work, new and affordable	Percent of low-income renter-occupied housing that is affordable	Percent of low-income owner-occupied housing that is affordable	Percent of low-income housing that is affordable
NJ	7.84%	0.82%	39.96%	1.51%	37.51%	35.32%	36.27%
MD	8.23%	1.41%	38.31%	2.19%	44.97%	39.76%	41.82%
WV	14.10%	0.02%	40.44%	3.30%	45.74%	52.38%	50.82

Table 5: Comparison of State wise proportions

There are a larger percentage of single mothers living in unaffordable rental units, in NJ (57.5%), than in the other two states. For owners, the difference between NJ and MD is very small and higher than WV. This is also true for seniors: 51.31% of owners and 66.32% of renters in NJ live in unaffordable housing. Overall, WV has the lowest percentage of single mothers and seniors living in unaffordable housing. Figure 12 represents these differences.

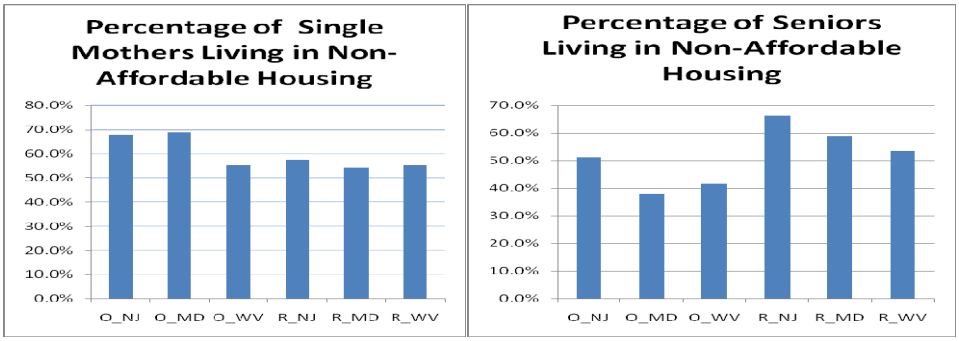


Figure 7: Percentage of single mothers (left) and seniors (right) living in non-affordable housing.

4. Regression Analysis.

a. Variable Selection

We compared the three states separately for owner-occupied and renter-occupied residences, which resulted in analyses of six subsets of the data. The different models selected for each of six subsets are in the table below. Where several variables that represent similar demographic information were important, we decided to keep those variables that tend to zero more slowly than the others, as shown by coefficients plots from LASSO and elastic net. The following tables below represent the result from logistic regression estimation for owner occupied households and rental units for all three states. By interpreting the coefficients in these tables, you can see whether a change in the variable predicts a decrease or increase in the probability of the housing being affordable.

	Coefficients	SE	Wald Chi Square	Odds Ratio
INTERCEPT	-25.8181***	0.4286	3628.17	
SENIOR	1.2792***	0.0314	1664.69	3.594
ROOMS	-0.3096***	0.0080	1510.44	0.734
LHINC	2.6058***	0.0425	3751.09	13.543
PERSONS	-0.2117***	0.0107	390.96	0.809
TRAVEL_TIME	0.2063***	0.0282	53.62	1.229

Table 6: Logistic regression estimates of owner occupied households in New Jersey
*Response variable: Affordability (0) Not affordable (1) Affordable
*Significant variables from stepwise regression: sex, educoll, senior, persons, rooms, lhinc

	Coefficients	SE	Wald Chi Square	Odds Ratio
INTERCEPT	-13.7926***	0.3364	16980.99	
SENIOR	1.5975***	0.0292	2986.67	4.941
ROOMS	-0.2528***	0.0071	1261.89	0.777
EDUC	-0.1119***	0.0048	555.72	0.894
LHINC	1.5075***	0.0335	2024.86	4.515

Table 7: Logistic regression estimates of owner occupied households in Maryland
*Response variable: Affordability (0) Not affordable (1) Affordable
*Significant variables from stepwise regression: educoll, educ, age, senior, rooms, lhinc

	Coefficients	SE	Wald Chi Square	Odds Ratio
INTERCEPT	-12.0305***	0.3138	1469.85	
SENIOR	0.6308***	0.0328	369.91	1.879
ROOMS	-0.1865***	0.00994	351.99	0.830
LHINC	1.2876***	0.0325	1565.85	3.624
BLDGSZ	0.3512***	0.0247	202.85	1.421

Table 8: Logistic regression estimates of owner occupied households in West Virginia
*Response variable: Affordability (0) Not affordable (1) Affordable
*Significant variables from stepwise regression: uncrowd, educoll, senior, bldgsz, rooms, lhinc

	Coefficients	SE	Wald Chi Square	Odds Ratio
INTERCEPT	-9.5621***	0.2839	1134.07	
LHINC	1.0090***	0.0283	1267.48	2.743
EDUCOLL	-1.0075***	0.0343	864.19	0.365
POVG	-0.0774***	0.0315	6.0359	0.926
PERSONS	0.1693	0.0104	265.09	1.185
ROOMS	-0.2428	0.0095	659.10	0.784

Table 9: Logistic regression estimates of rental units in Maryland
*Response variable: Affordability (0) Not affordable (1) Affordable
*Significant variables from stepwise regression: eduhs, educoll, persons, rooms, lhinc, hinch, renth

	Coefficients	SE	Wald Chi Square	Odds Ratio
INTERCEPT	-14.498***	0.2492	3385.82	
LHINC	1.5749***	0.0264	3571.52	4.830
EDUC	-0.1092***	0.0040	742.92	0.897
ROOMS	-0.2193***	0.0087	637.75	0.803

Table 10: Logistic regression estimates of rental units in New Jersey
*Response variable: Affordability (0) Not affordable (1) Affordable
*Significant variables from stepwise regression: smother, educoll, eduhs, sex, rooms, persons, lhinc, hinch, renth

	Coefficients	SE	Wald Chi Square	Odds Ratio
INTERCEPT	-8.4851***	0.5992	200.496	
ROOMS	-0.2785***	0.0226	151.70	0.757
LHINC	1.0470***	0.0607	297.68	2.849
POVG	-0.80027***	0.0688	136.14	0.448
BLDGSZ	0.0291***	0.0102	8.079	1.029

Table 11: Logistic regression estimates of rental units in West Virginia
*Response variable: Affordability (0) Not affordable (1) Affordable
*Significant variables from stepwise regression: bldgsz, hinch, lhich, renth

The final selection of all predictors to explain housing affordability has shown in table below. Some of predictors are only significant for owners (like travel time to work, head of household is a senior), some of predictors are significant for renters only (head of household is a college graduate).

	Owners			Renters		
Variable	NJ	MD	WV	NJ	MD	WV
Income	✓	✓	✓	✓	✓	✓
Number of rooms in the unit	✓	✓	✓	✓	✓	✓
Number of people in the household	✓				✓	
Highest level of education completed		✓		✓		
Head-of-household is a college graduate					✓	
Travel time to work is less than the median	✓					
Head-of-household is a senior citizen	✓	✓	✓			
Building size			✓			✓

Table 12: Final Predictors selected to explain affordability

General conclusions about the positive or negative effects of these variables on the likelihood of affordability could be inferred from these tables.

b. Results of variable selection methods – Owner-occupied units

For owner-occupied units in all states, whether the head-of-household is a senior citizen is a significant predictor of affordability: the odds of having affordable housing if a household is low-income, increases if the head-of-household is a senior citizen. For the owners subset, MD and WV differed from NJ in the following way: travel time to work is not a significant predictor of affordability for MD and WV, but for NJ, travel time to work being "good" – compared to the median for that state – was a significant predictor of affordability. The closer the head-of-household was to work, the larger likelihood of affordability of low-income housing. Also, the number of persons in the household contributed to the prediction of affordability in NJ (but not for MD or WV): an increase in the number of persons leads to a decrease in the odds of affordability. Using LASSO

and elastic net, the education level of the head-of-household and building size was important for MD and WV, respectively, but not for other states.
Descriptive statistics enabled identification of the differences between the states with regard to the number of persons in the household: amongst owners in NJ, only 26.7% are householders living alone. However, out of this 26.7%, 64.9% live in unaffordable housing and 73.6% are females. Only 25.1% of females living alone in NJ have affordable housing. In MD, only 31.0% of females living alone in owner-occupied, low-income housing have affordable housing, compared to 26.3% in WV. Also, NJ has the lowest percentage of renters females living alone 22.57% compared to MD 26.9 % and WV 29.6% (Figure 8).

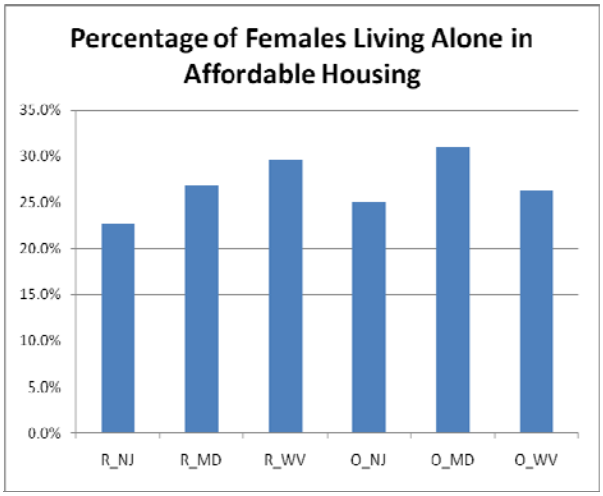


Figure 8: Percentage of Females living alone in affordable housing.

The results from the stepwise regression confirm that the education, number of rooms in the household, and the senior status of the head-of-household are important predictors of affordability. Again, building size was important in predicting affordability in WV but not in other states. For the rooms variable, we created new dummy variables for each number of rooms to see whether a binary variable would be more important than the nominal variable; this was true only for a binary variable to indicate that rooms = 7, for owners in NJ. The following graph presents the distribution of rooms for owners in NJ.

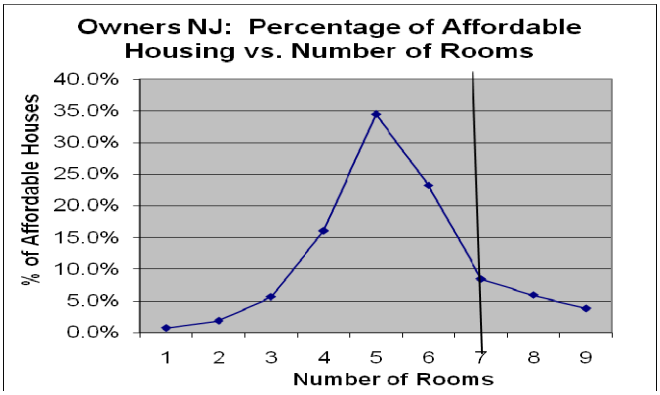


Figure 9: NJ Owners - Percentage of affordable houses vs. number of rooms in the household.

It is also interesting to note that for WV, education was not selected by the LASSO and elastic net procedures, although it was selected by the stepwise procedure. From this, we can generalize that education is an important predictor of affordability for owner-occupied housing. Note that in WV, the building size was the last variable selected by the method. Almost all of the owner-occupied, low-income units in WV are mobile homes, as seen in Figure 10.

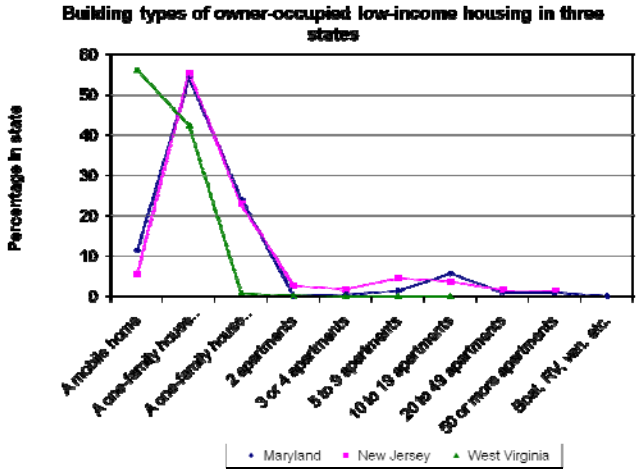
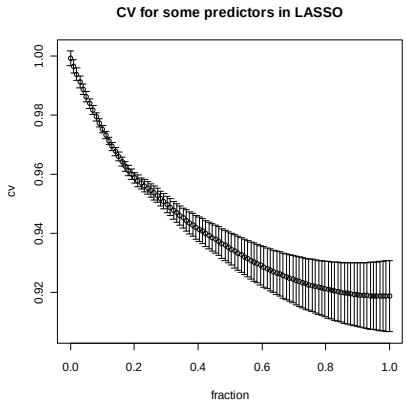


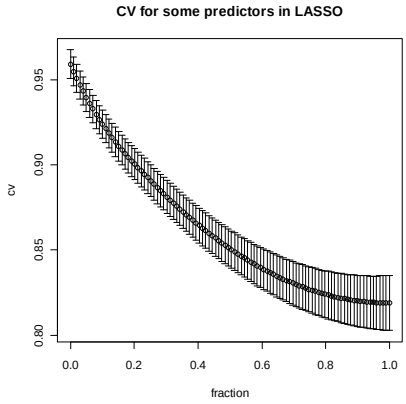
Figure 10: Distribution on building types for owner-occupied low income housing.

When this model was refitted using a dummy variable to represent mobile homes, this new variable was important but also on the borderline for being selected.

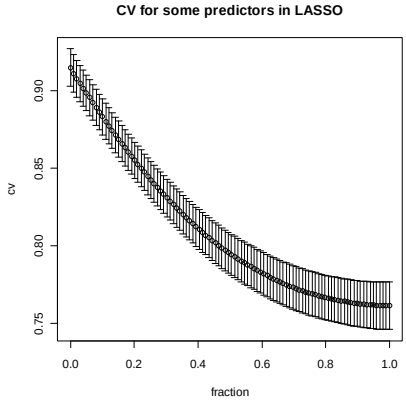
Finally, we verified our final model using a cross-validation error plot from a new LASSO model, using only our variable selections listed in Table 12 in the text. As shown in Figures 11, there are no variables in the model that we should consider removing.



Owner-occupied, low-income housing in West Virginia



Owner-occupied, low-income housing in Maryland



Owner-occupied, low-income housing in New Jersey

Figure 11: Cross-validation error plots for final subsets of variables, For owner-occupied, low-income housing

c. Results of variable selection methods – Renter-occupied units

For rental units, when judged using LASSO and elastic net, education and poverty status were important for at least two of the three states. There are a larger proportion of households below the poverty line in WV and MD, than in NJ. On average, the probability of a rental unit being affordable, given that the household is below the poverty line, is lower than the probability if the household was above the poverty line, for WV and MD. There is no significant effect on affordability for NJ due to poverty status.

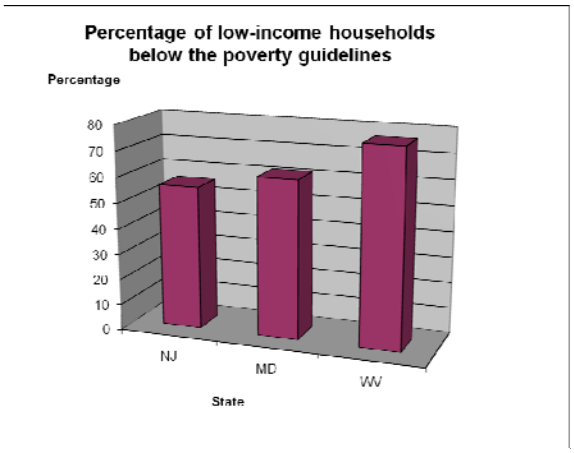


Figure 12: Percentage of Renters low income households below the poverty guidelines.

Education is an important predictor, for NJ and MD (both educational level and college) but not for WV. For parsimony, we decided to keep the education variable that tends to zero slower than the other, as described above in our discussion of the renters in MD. This leads to our selections in Table 12 here in the text. With regard to the number of persons in the household, for renters, we again found that a low percentage of females living alone have affordable housing.

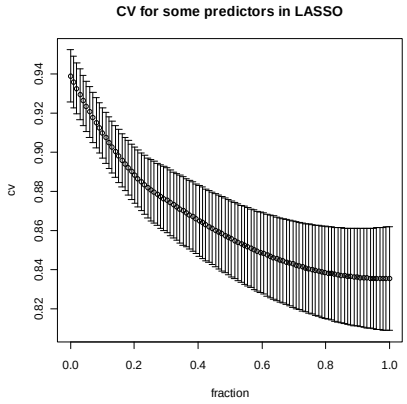
In NJ this percentage was 22.6%; MD, 26.9 %; and WV, 29.8%. Again, for WV, building size was the last variable chosen by these methods, although it was not selected for other states. When the model was re-fitted using a binary variable to represent mobile homes, this new variable was more important than most of the other variables, but not important enough to be kept in the final model. The same was true for a dummy variable representing apartment units.

For the rental units in WV, it is interesting to see that only the building size was a common variable selected by these two selection methods (the income was as well, but this is was expected). Also, it was only important for WV, not for the other two states.

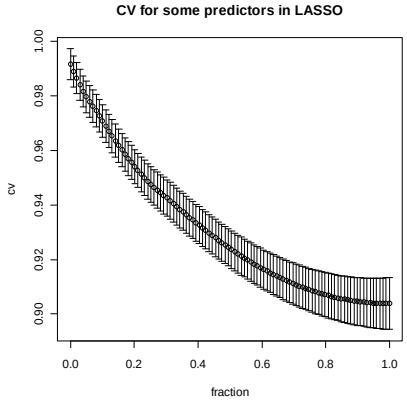
For the rental units in MD, the variables representing college education, persons, rooms, and income were also included in a stepwise model selected by the Akaike's Information Criterion (AIC) for each subset.

For NJ, the variable college education was in a model from stepwise, while we chose to use the highest level of education for our final model through elastic net. The overall observation here is that education is definitely an important variable to predict affordability of rental housing in NJ and in MD, with more education decreasing the odds of obtaining affordable housing for low-income households. Combined with the observations from the owner-occupied housing, we can see that education is probably important for predicting affordability of all housing in these three states; several different variable selection methods had these variables in the final selections.

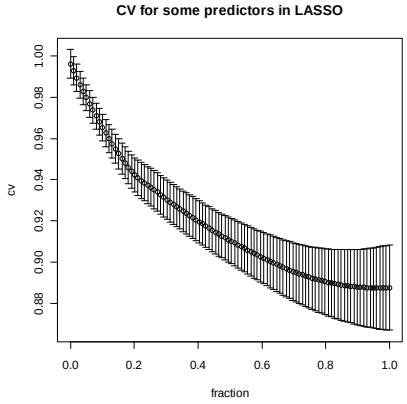
Note that in all six subsets show a similarity in the education variable. Most of the householders in these low-income households only hold a high school diploma. As shown in Figure 13, the cross-validation error doesn't stop decreasing significantly until lambda = 1. This confirms that our model has no extraneous variables.



Renter-occupied, low-income housing in New Jersey



Renter-occupied, low-income housing in Maryland



Renter-occupied, low-income housing in West Virginia

Figure 13: Cross-validation error plots for final subsets of variables, For renter-occupied, low-income housing

III. Conclusion

The data seems to provide evidence that there are important differences between income levels and housing costs, types, and affordability, between New Jersey (NJ), Maryland (MD), and West Virginia (WV). Income levels tend to be highest in NJ, and lowest in WV. WV and MD have a higher percentage of low income households living in mobile homes, than the other two states. However, WV and MD have higher percentages of affordable low-income housing than NJ. WV has the lowest percentage of single mothers and NJ has the highest percentage of senior citizens (62+) living in unaffordable housing, than the other states.

Overall, the educational level of the head-of-household is an important predictor of the likelihood of affordability for housing units in NJ, MD, and WV. There is some evidence that an increase in educational level leads to a decrease in the likelihood of affordability, for low-income households. Whether this is true for all households, is outside the scope of this report.

The poverty status of households matters more in MD and WV in predicting the probability of living in affordable housing, than it matters in NJ, but again, there is more affordable low-income housing in WV and MD than in NJ. Also, building types for those who live in low-income housing differs significantly in WV from building types in NJ and MD. Specifically WV has more mobile homes and fewer one-family homes than NJ and MD, and for this state, the type of the building may have an affect on the odds of affordability, but this requires further investigation.