5-2020

An Econometric Analysis of Homelessness Risk-Factors

Zac Smith

An Econometric Analysis of Homelessness Risk-Factors

by

Zachary W. Smith

Advisor: Dr. Abel Embaye

An Honors Thesis in partial fulfillment of the requirements for the degree Bachelor of Science in Business Administration in Economics.

Sam M. Walton College of Business
University of Arkansas
Fayetteville, Arkansas

May 11, 2020

Abstract

This paper explores the issue of homelessness within the United States and seeks to create an econometric model that identifies predicting factors of homelessness at a state level which can be used to estimate the size of homeless populations. The author analyzed the role of 16 factors including income inequality, minimum wage, unemployment, rental cost, poverty, education, veteran status, and substance abuse on the 2017 state homeless populations. Using an ordinary least squares regression, the model produced five significant variables: adjusted minimum wage, percent of income spent on rent, possession of health insurance, average winter temperature, and the unemployment rate. Through its assessment of economic, personal, and environmental factors, this model provides a foundational understanding of the types of variables which predict state homelessness.

Introduction

Experiences of homelessness are all too common in the United States; the homeless are commonly seen in parks, alleys, and at sitting on street corners. In fact, on one night alone in 2017, over 553,000 individuals were experiencing homelessness, a third of which were unsheltered (Watt, Henry, Rosenthal, Shivji, & Abt Associates, 2017, p. 1). Although there are demographic trends that can be found within homeless populations, homelessness is no respecter of race, ethnicity, or religion. Whether by President Reagan implying that homelessness is a choice-driven lifestyle or by stereotypes driving the public understanding of the problem, the plight of this vulnerable group is often oversimplified and misrepresented (Parsell & Parsell, 2012).

Homelessness is an important social ill to study not only because it indicates a failure of society to care for its most vulnerable group, but also because the homeless present society with an enormous economic cost. The United States Interagency Council on Homelessness has backed policies such as supportive housing initiatives due to the fact that ending chronic homelessness can help decrease social costs associated with emergency care, jail, and public homeless shelters (2015). Some research indicates that roughly one-third of emergency room (ER) visits are made by the chronically homeless as the average homeless person make 5 trips to the ER a year; this leads to an average annual cost of $18,500 per homeless individual. This is a very costly amount, especially considering that 80% of the trips concern problems which preventative care could have taken care of (Garrett, 2012). Other sources such as the National Alliance to End Homelessness states that the total average cost a persistently homeless individual places on taxpayers is $35,578 (2015, p. 1). Regardless of the exact individual cost, the U.S. Government allocated close to $5 billion in 2016 to support the cause of homelessness (U.S. Interagency Council on Homelessness, 2016). This problem is far more than just social in nature as its implications are heavily economic.

In order to address an issue like homelessness, policy makers and social activists alike must understand the causes of homelessness, whether it is primarily economically motivated, whether it is simply due to mental illness, or whether it is a collection of factors. This study attempts to create a predictive model through multiple regression using ordinary least squares. To my best knowledge, this work is one of the first academic papers to conduct an econometric analysis on recent homelessness data while incorporating variables and factors which span more than just a purely demographic or economic perspective. As such, this study does not use variables which are new to the study of homelessness, but rather aggregates factors which have been discussed in a variety of academic/professional works and regresses them using recent homelessness data.

The purpose of this study is to find statistically significant factors in a state populace that can predict rates of homelessness which in turn can support researchers in understanding of what elements are either risk-factors or indicators of homelessness. While assessing 16 different independent variables, the final model includes 5 statistically significant variables which explain approximately 78.72% of the variance in the state homelessness rates. While this model does not provide proof of a causal relationship between any given variable and rates of homelessness, it can at the least help provide additional insight into predicting rates of homelessness without relying on costly head-counts as well as provide additional understanding of the dynamics which drive this phenomenon.

Literature Review

This seemingly innocuous question has far-reaching effects on how homelessness is measured and addressed at a governmental level. At a surface level, defining homelessness appears to be a straightforward question, however, creating a consistent definition that allows for accurate data collection and provides quantitative insight is quite difficult. Between various government departments, homelessness is defined a myriad of ways. In fact, the U.S. Government Accountability Office created a report bringing out this very point and urging various agencies to develop more consistent terminology. They reported that while the Departments of Education, Health and Human Services, and Housing and Urban Development all conduct surveys and provide data sets on various topics concerning homelessness, their methods and definitions were not congruous. As a result, it is quite complicated if not impossible to combine the datasets to give a clearer and more accurate, multifaceted understanding of homelessness. Furthermore, this lack of cohesive terminology affects the homeless themselves as their eligibility for various services can vary drastically are create a great deal of confusion resulting in families and individuals either being underserved or unaware of available resources (Government Accountability Office, 2010).

Despite this clear acknowledgement of definitional inconsistency, departments still have varying terminology close to a decade later. For example, the U.S. Department of Housing and Urban Development (HUD) had four classifications for homelessness in 2018. Those categories included the literally homeless, those in imminent risk of homelessness, those that are homeless under other federal statutes, and those who are fleeing or attempting to flee domestic violence (United States Interagency Council on Homelessness, 2018). On the other hand, the McKinney-Vento Homeless Assistance Act which was passed 1987 in order to provide the Homeless with improved federal benefits, includes in its definition of homeless children and youth, those which are sharing housing or doubling up due to, "loss of housing, economic hardship, or a similar reason" (42 U.S.C. § 11434a(2)). The discrepancy in this situation is that when HUD reports annual homelessness figures to Congress each year, people that are doubling up and sharing homes are not counted as homeless. This presents a clear shortcoming when taking into account research that took place after the 2008 recession. In a report prepared for HUD providing an analysis of trends in household composition, researchers found that the financial crisis before the recession, as well as the recession itself, contributed to a constant decrease in household formations from 2003 to 2009 as well as the increased number of unrelated families living in one home (Eggers & Moumen, 2012, pp. 40). Another article from 2018 reports that households containing more than one family have more than tripled in the past 14 years (Bush & Shinn, pp. 1). This simple fact indicates that while many of the children in these household could classify as homeless under the McKinney-Vento Homeless Assistance Act, since they are not imminently homeless, they are not included in HUD's annual report to Congress. Consequentially, the number of people reported to be experiencing homelessness is minimized. This example just goes to show why each data set and piece of information about homelessness must be carefully assessed in order to document accurate trends within the homeless population and provide coherent results.

Although differences in terminology and data collection methodology can and do impact the way data must be handled when using information from multiple government agencies, HUD has one of the best sources of compiled data which estimate the number of literally homeless individuals in the US. As such, HUD's terminology and definitions will be of most importance to this specific study. According to the 2018 Annual Homelessness Assessment Report published by HUD, "Homeless describes a person who lacks a fixed, regular, and adequate nighttime residence" (Henry et al., 2018). The document then proceeds to state that chronically homeless individuals are people with a disability which have either continuously experienced homeless for a year or more or which have had four or more experiences of homelessness within a three year time-span where the total time homeless is at least 12 months. Within this homeless population, there are those which experience sheltered and unsheltered homelessness. Sheltered homelessness refers to those which stay in "emergency shelters, transitional housing programs, or safe havens" and unsheltered homelessness refers to people who primarily stay in locations not suitable for sleeping such as in parks, cars, or on streets (pp. 2-3). These types of homelessness can then be broken down by various factors to address homeless individuals, homeless families with children, unaccompanied homeless youth, etc. Understanding and defining homelessness is the first step to be able to properly address its issues and analyze its causes, however, the way in which these definitions are measured and turned into metrics is an important process which also greatly impacts the way in which a study must be conducted.

Measurement Techniques

In 2001 Congress directed HUD to enumerate the homeless and conduct an annual assessment in order to track the progress of the McKinney-Vento Homeless Act (Committee on an Evaluation of Permanent Supportive Housing Programs for Homeless Individuals [CEPSHP], et al., 2018). For this data, HUD turned to the local Continuum of Care (CoC) bodies. The CoC Program was set out in Subtitle C of Title IV of the McKinney-Vento Homeless Act which allowed for the federal support of regional bodies which connect the local homeless with access to State and local resources in order to promote self-sufficiency of the homeless (42 U.S.C. § 11381-11389). In addition to providing and channeling resources, each year, the CoCs are also expected to provide a head count and survey data of the local homeless population. HUD then aggregates and analyzes this data before releasing it in the Annual Homelessness Assessment Report (AHAR) to Congress in order to provide an update on the state of domestic homelessness. HUD also worked to develop the Homeless Management Information System (HMIS) which helps CoCs collect and store demographic and other personal data on their clientele (CEPSHP, 2018).

Each AHAR has two parts, although the second part typically takes an extra year to create and publish, and they both provide demographic data and insight into how the homeless use emergency housing resources. The first part of the AHAR includes a point-in-time count (PIT count), which provides a stock variable describing how many homeless individuals there are in one night in January, as well as a housing inventory count (HIC count) which sums up the total number of beds and units dedicated to support the homeless (U.S. Department for Housing and Urban Development [HUD], 2020). The second part of the AHAR provides a flow variable which estimates how many people annually access and utilize resources provided by CoC.

In order to the conduct the PIT count during the last 10 days of January, CoC's have a few tools they use. For the sheltered homeless count, the HUD *PIT Count Methodology Guide* suggests four main methods: complete census count, random sample and extrapolation, non-random sample and extrapolation, or a combination of census and sampling. While HUD prefers using either a census or random sample due to its more accurate representation of the greater populace, these are also very resource intensive methods. CoC's often settle for non-random sample and extrapolation where they utilize HMIS data from a single homeless shelter and expand it to model what they believe to be accurate (HUD, 2014, pp. 15). Conducting a PIT count of the unsheltered homeless, on the other hand, is quite a bit different. Since the unsheltered are much more difficult to access and count, CoC's generally have a "night of the count" where from dusk to dawn on one designated night, surveyors go out with flashlights and a clipboard and count all of the unsheltered homeless individuals they see. There are a few different ways this count is implemented. Some CoC's use a complete coverage count where the whole geographic area is canvassed in a fairly equal manner while others use a known location count where enumerators count all of the homeless in specific regions that are know to host unsheltered homeless. A random sample of areas count is also used to support the known location count method as it provides additional accuracy and finally, a service-based count sometimes occurs within 7 days after the night of the count. This allows surveyors to interview people at various public and private locations and determine whether they are an unsheltered, previously uncounted homeless individual (Department of Housing and Urban Development, pp. 18-19).

The second part of each AHAR publication includes an annual estimate of how many homeless people use an emergency shelter, transitional housing program, or permanent supportive housing program. This data is much easier to collect as HUD primarily aggregates the information from over 375 CoCs nationwide. While this count is unduplicated, it does represent a sample which is adjusted to represent the population as a whole. This survey provides a better understanding for how many individuals experience homelessness through the year rather than assuming the year-long figure is simply an extension of the PIT count. One primary drawback, however, is that this estimation does not include the unsheltered homeless (Office of Community Planning and Development [OCPD], 2018, pp. x).

Each of these methods has its own set of rules and regulations to ensure maximum statistical accuracy and minimize sampling bias, however, due to the nature of this population, it is very difficult to produce a data set with accurate representation and no duplicity. While counting the sheltered homeless is somewhat simplified due to the more formal process and presence of the HMIS, counting the unsheltered is more problematic and prone to error. One study which delved into unsheltered homeless enumeration techniques said that even well-executed street counts could miss more than half of the homeless due to various enumerator biases (Hopper, Shinn, Laska, Meisner, & Wanderling, 2008). Despite the difficulty in providing an error-free estimate of the homeless population, HUD provides some of the most consistent data available on the issue and keeps its limitations in mind while drafting its reports.

Causes of homelessness

Although there are many factors which contribute to homelessness, academic and professional literature often target similar components. In a recent article drafted within the Executive Office entitled, "The State of Homelessness in America", the authors attributed homelessness to a few core factors: higher price of housing due to overregulation of housing markets, the tolerability of sleeping outside, supply of homeless shelters, and certain characteristics of individuals/communities such as incarceration, low income, substance abuse, and severe mental illness (Council of Economic Advisors, 2019, pp. 1-6). Another study entitled, "The Ecology of Homelessness" takes a more generalized look at all of the possible factors contributing to homelessness and proposes a fairly comprehensive ecological model (see Appendix A) which examines the dynamic relationship between the following four factors: biopsychosocial risk factors, individual and social outcomes, the temporal dimension, and housing outcomes (Nooe & Patterson, 2010, pp. 106-107). The National Coalition for the Homeless posits that poverty and a shortage of affordable rental housing largely affect homelessness as well as other factors including a lack of affordable health care, domestic violence, mental illness, and addiction disorders. While there are an endless number of micro factors which could lead one down a path resulting in homelessness, research in the field seems to largely revolve around housing, poverty, and individual factors.

Affordable housing is often one of the first areas considered when studying factors which contribute to homelessness, and with reason, for much research supports this idea. In a journal article examining this very point, the authors concluded that, "…the rental vacancy rate exerts a negative and statistically significant effect on homelessness, while measures of housing costs such as median rents and rent-to-income ratios exert positive and significant effects" (Quigley & Raphael, 2001, pp. 333-334). As such, housing factors are important to consider when examining the issue of homelessness in the United States. Between 2005 and 2015, around 34 million families and individuals began living in rental housing due to the housing bubble, decreased incomes, and risk aversion. This increase in demand was met with an increase supply as 8.2 million units were added to the market, however, low-income households were at risk. Of the newly-constructed units, only 10% had price-levels below $850, a rate which approximately half the renters could afford according to the 30-percent-of-income standard (Joint Center for Housing Studies of Harvard University [JCHSHU], 2015, pp. 1-3). This market can be seen further constricting through the rental vacancy rate which has dropped significantly since 2009 as it reaches a 35-year low vacancy rate of 6.4% in the fourth quarter of 2019 (U.S. Census Bureau, 2019a). In addition to the problems that inadequate housing brings, one study found that households that spent over 50% of their income on housing spend 38% less on food and 55% less on healthcare (JCHSHU, pp. 5). From this it is clear that affordable housing presents a real and present problem to the homeless and those in low-income situations.

Another primary risk factor which people attribute homelessness to is poverty. Although poverty and homelessness have a bi-directional relationship, much pre-existing research finds that homelessness is most often precursed by the experience of poverty (Johnsen & Watts, 2014, p. 37). When assessing the ways in which poverty contributes to homelessness, the National Coalition for the Homeless point to decreasing employment opportunities, declining wages, wealth inequality, and declining public assistance (2007, pp. 1). Poverty is also pointed to in relation to its impact on affordable housing as increasing rates of poverty within the lower-class leads to an increase in demand for low-income housing and raises the price above that which the most impoverished can afford (Wright & Rubin, 1991). Another article which analyzed employment and earnings of the homeless in New York City expressed that job loss is often a triggering factor which contributes to a cycle of homelessness. Furthermore, it found that the adults with families had an employment rate of 42% before homelessness and the working individuals made an average of $8,483 compared to the 2008 poverty line of $14,000 for a household of two (Metraux, Fargo, Eng, & Culhane, 2018, pp. 183). Poverty is a key element in addressing homelessness as without gainful-employment and sufficient income, families and individuals can be quickly evicted and cast into the unstable environment of homelessness which further exacerbates the conditions they are fighting.

Along with the issues of affordable housing and poverty there are many individual characteristics which are also associated with homelessness. Two of the most prominent of these characteristics being severe mental illness and substance-abuse/addiction (Fowler, Hovmand, Marcal, & Das, 2019, p. 467). Some statistics state that between 20 and 30% of the adult homeless have a severe mental illness and around half of the single adults have either in the past or presently struggled with alcohol or drug addiction (Foscarinis, 1996, pp. 6-7). In fact, the 1980's theories of de-institutionalization and abandonment, which essentially state that de-institutionalization of mental hospitals led to patients being abandoned by their care provider and resulted in a surge of homelessness, have played a significant role in shaping the academic narrative of homelessness (Institute of Medicine (US) Committee on Health Care for Homeless People, 1988). Another factor commonly related to homelessness is social exclusion and the fact that minority racial and ethnic groups are at greater risk to become homeless than majority groups. This can be due to income inequality, differences in educational attainment, and the higher incarceration rates which groups like African-Americans face at a rate disproportionate to that of their white counterparts (Shinn, 2010, pp. 29-35). To exemplify this, a report prepared for the Abdul Latif Jameel Poverty Action Lab shared that African Americans represent 41% of the homeless population although they are a mere 13% of the US population. Other factors which this report brings out are that veterans are 25% more likely to experience homelessness than non-veterans and that while those fleeing domestic violence are 5% of the population, they make up about 17% of the homeless population (Evans, Phillips, & Ruffini, 2019, pp. 6). While the micro-factors which can increase individual's risk of becoming homeless are countless, it is necessary to consider the homeless population's intersectional identities in order to better understand the specific factors which can lead to homelessness.

Model and Data

For the initial model, all of the variables will be included in order to create the following multivariable equation:

Homeless = β_0 + β_1(Gini) + β_2(MinWage*CoL) + β_3(Income*CoL) + β_4(RentCost*CoL) + β_5(RentVac) + β_6(Poverty) + β_7(Unemp) + β_8(Veteran) + β_9(Education) + β_{10}(Insurance) + β_{11}(Disability) + β_{12}(Mental) + β_{13}(Alcohol) + β_{14}(Cocaine) + β_{15}(Heroine) + β_{16}(Temp) + ε

These beta-values will then be estimated through a least-squares linear regression. In terms of hypothesis testing:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_{16} = 0$$
$$H_1: At\ least\ one\ \beta_i \neq 0$$

Finally, as our sample size, *n* is 50, and the number of parameters in our equation is 17, when referencing the T-statistic, the total degrees of freedom will be 33. All test statistics will be compared to $\alpha = 0.05$. From this initial equation and set of hypotheses, the model will be regressed and tested utilizing *EViews 10 Student Version Lite* and then modified accordingly to meet the core assumptions of multiple regression.

In order to create a model that predicts homelessness across the 50 US states, prior research was relied upon to provide guidance in the variable selection process. The name, description, and source of all variables used in this study can be found in Appendix B and the variables' descriptive statistics can be found in Table 1. The dependent variable was selected to be the state point-in-time homeless count from 2017, so all the independent variables were selected from the years 2014-2019 in order to give an accurate depiction of what external factors in 2017 were like. To normalize dollar values across states, the Cost of Living index was divided by 100 to yield a decimal which was then multiplied by the Minimum Wage, Income, and Rent Cost variables in order to provide composite variables which take into account varying rates of purchasing power between states.

The variables included were selected in order to introduce factors from different facets of of an individual's life which might contribute to homelessness. The Gini coefficient for example was selected to test whether a state's level of income inequality contributes to homelessness. Factors like the minimum wage, income, cost of rent, rental vacancy rates, poverty rates, and unemployment rates help test various economic factors which can greatly affect one's quality of life. Other explanatory variables like veteran status, levels of educational attainment, possession of insurance, disability rates, mental illness rates, and rates of various substance abuse help test whether personal factors which are measured at a state-level could contribute to the state's homelessness rate. And finally, temperature is included since the point-in-time indicator is taken in January and some research indicates that natural environmental conditions impact homelessness rates.

The data is primarily from government sources due to its reliable nature, however one source of potential error in this study is the research methods and error within the datasets themselves. Another shortcoming of this study is that it relies on cross-sectional data and attempts to predict state homelessness with state-level explanatory variables. This could be problematic since states have so much internal variance of homeless population density. Ideally, longitudinal data would be used to regress the predictors of homelessness at an individual level and then incorporate more macro-economic factors, however, individual data sets are widely unavailable due to privacy concerns and data for several of the independent variables chosen below are not published on an annual basis. Another source of potential error is the fact that the dependent variable includes the chronically and temporarily homeless, sheltered and unsheltered alike. This could muddle the regressors as there may be stronger relationships between sub-groups of the homeless and these variables than between the overall group and the variables.

Variable	Mean	Median	High	Low	Standard Deviation	Sample Size
Homeless	14.2816798	10.6578781	46.9084366	4.522248829	9.666033359	50
CoL	104.52	96.55	201.3	84.5	20.26955352	50
Gini	0.464706	0.4663	0.5157	0.4225	0.018838433	50
MinWage	8.2278	8.15	11	5.15	1.275824894	50
Income	22580.06	21783	31262	15127	3860.210504	50
RentCost	956.18	877	1573	690	207.1713001	50
RentVac	7.49	7.4	16.8	3	2.946676093	50
Poverty	12.232	11.85	20.8	6.5	2.87446969	50
Unemp	5.04	5.2	7.6	2.9	1.047091209	50
Veteran	6.7732	6.965	9.29	3.96	1.204131953	50
Education	38.938	38.15	54.1	29.7	5.016129584	50
Insurance	8.156	8	17.3	2.8	3.005073044	50
Disability	13.312	13.05	20.2	9.6	2.24514053	50
Mental	0.04371176	0.04364225	0.0552221	0.033189107	0.005290864	50
Alcohol	0.06064847	0.05952688	0.07476646	0.042995925	0.007968613	50
Cocaine	0.01708566	0.01604615	0.03052419	0.009735888	0.005271495	50
Heroine	0.00370127	0.00312841	0.01002546	0.001288095	0.00184897	50
Temp	32.228	32.05	67.4	2.6	12.21189649	50

Table 1: Descriptive Statistics

Results

After regressing the initial model, the results in Figure 1 were attained. From the initial equation an adjusted R-squared of 82.62% was achieved and the F-statistic of 15.55 is highly significant at $\alpha = 0.05$. Only five of the 17 parameters, however, were significant: MinWage*CoL, Income*CoL, RentCost*CoL, Insurance, and Temp.

Before simply removing all variables, which are not statistically significant, it is necessary to assess whether multicollinearity could be affecting the individual significance of the variables. To check this, the variance inflation factors (VIF) can be analyzed. While there is no hard rule on what VIF value indicates multicollinearity, a VIF value over 10 is generally an indicator that multicollinearity is impacting the regression coefficients (Akinwande, Dikko, & Samson, 2015). In Figure 2 the centered VIF values can be seen and it is immediately obvious that MinWage*CoL and Income*CoL need to be addressed as they both have values greater than 50. To support the idea that these composite variables are dealing with excessive correlation, a correlation matrix can be calculated found in Appendix C. From this figure and the VIF indication, it is clear that MinWage*Col and Income*Col are introducing multicollinearity into the model as their correlation is 0.827. Furthermore, the correlation between MinWage*CoL and RentCost*Col is equal to 0.878 and the correlation between Income*CoL and RentCost*CoL is 0.930. While this correlation is intuitive, it must be reduced, or a variable must be removed in order to decrease the effect of multicollinearity within the model.

Dependent Variable: HOMELESS
Method: Least Squares
Date: 03/23/20 Time: 12:27
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.348314	45.21289	0.051939	0.9589
GINI	-13.22677	76.20954	-0.173558	0.8633
MINWAGE*COL	1.958110	0.705494	2.775515	0.0090
INCOME*COL	-0.001812	0.000521	-3.479406	0.0014
RENTCOST*COL	0.042593	0.009091	4.685004	0.0000
RENTVAC	0.209756	0.347893	0.602933	0.5507
POVERTY	0.074173	0.507725	0.146089	0.8847
UNEMP	-1.485618	1.080941	-1.374375	0.1786
VETERAN	0.950802	0.792015	1.200484	0.2385
EDUCATION	0.270862	0.255067	1.061923	0.2960
INSURANCE	0.642826	0.309446	2.077346	0.0456
DISABILITY	-0.759895	0.717545	-1.059022	0.2973
MENTAL	311.9993	218.6031	1.427241	0.1629
ALCOHOL	-119.2706	126.6199	-0.941957	0.3531
COCAINE	137.6017	261.5192	0.526163	0.6023
HEROINE	-480.8466	626.8833	-0.767043	0.4485
TEMP	-0.420702	0.096129	-4.376450	0.0001

R-squared	0.882923	Mean dependent var	14.28168
Adjusted R-squared	0.826159	S.D. dependent var	9.764168
S.E. of regression	4.071095	Akaike info criterion	5.910186
Sum squared resid	546.9360	Schwarz criterion	6.560274
Log likelihood	-130.7546	Hannan-Quinn criter.	6.157743
F-statistic	15.55418	Durbin-Watson stat	2.040704
Prob(F-statistic)	0.000000		

Figure 1

Variance Inflation Factors
Date: 03/23/20 Time: 16:26
Sample: 1 50
Included observations: 50

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	2044.205	6166.971	NA
GINI	5807.893	3789.970	6.218070
MINWAGE*COL	0.497722	126.3904	11.63652
INCOME*COL	2.71E-07	528.0925	55.31606
RENTCOST*COL	8.27E-05	317.2489	51.19376
RENTVAC	0.121030	23.65381	3.170332
POVERTY	0.257785	122.7847	6.425702
UNEMP	1.168433	93.40374	3.864739
VETERAN	0.627288	89.56037	2.743864
EDUCATION	0.065059	302.5192	4.938499
INSURANCE	0.095757	21.82506	2.608714
DISABILITY	0.514870	283.0824	7.829471
MENTAL	47787.31	279.4945	4.035647
ALCOHOL	16032.61	180.9775	3.071263
COCAINE	68392.28	65.96423	5.733531
HEROINE	392982.6	20.29440	4.053035
TEMP	0.009241	33.11209	4.157372

Figure 2

One way in which to do this, is to simply create another composite variable. Rather than including the composite variables of lowest quintile income and median cost of rent, both adjusted for cost of living, these can be combined to create a third composite variable measuring median cost of rent divided by lowest quintile income. In order to create more coherent results, this value is multiplied by 12 since RentCost is a monthly figure while Income is annual. This in effect provides an indicator for what percent of annual income is spent on housing by those with the bottom quintile of income. Once this variable is introduced and replaces its constituents, there is no VIF value in the model over 10 and the specific VIF value for the new composite variable is 7.828 while its correlation with MinWage*CoL is 0.463. The resulting regression coefficients are depicted in Figure 3. This new model is still highly significant with an F-statistic probability of 0.000, however, there are only four significant variables now: MinWage*CoL, RentCost/Income, and Temp.

In order to proceed with the removal of variables, a stepwise-backward algorithm is utilized. Although this algorithm has been critiqued in econometric analyses for sometimes choosing nuisance variables rather than true variables which leads to a lack of accuracy when extrapolating, this problem is more typical when used with big data and many predictive variables (Smith, 2018). This variable-removal methodology takes a model with all the desired variables included and then begins by removing the variable with the highest p-value. The resulting model then has its variable with the highest p-value also removed. Both of these variables are then compared to a cutoff value provided by the user and if either variable has a p-value lower than this number, the variable is reinstated. This process is then iterated until the largest p-value in the model is less than the cutoff value (IHS Global Inc., 2019). As this study has utilized an alpha level of 0.05, the cutoff value used in the stepwise-backward algorithm is also 0.05. The results from this are shown in Figure 4.

In this revised model, the variables have been trimmed down to MinWage*CoL, RentCost/Income*12, Insurance, Temp, and Unemp which are all significant. Between the adjusted R-squared value of 77.99% and the F-statistic p-value of 0.00, this model has relatively strong predictive power. In order to determine whether this is a final model, however, the assumptions of linear regression must be tested. Having reduced multicollinearity the next test to run is the Breusch-Pagan-Godfrey test which indicates whether there is hetroskedasticity in the estimation. The null hypothesis of this test is that error variances of all observation are equal while the alternate hypothesis states that not all error variances are equal (Breusch & Pagan, 1979). In other words, the null states that residuals are homoscedastic and the alternate states that residuals are not homoscedastic. The results from this test can be seen in Figure 5. As the image shows, the probability value for the observed R-squared value is 3.29% which is significant, hence rejecting the null hypothesis and indicating that the residuals are heteroscedastic.

In order to correct for heteroscedasticity, a natural logarithmic transformation is used on the Homeless variable. When the Breusch-Pagan-Godfrey test is rerun, the observed R-Squared value's probability value is now 56.35% as shown in Figure 6. This value is clearly insignificant, indicating that the test fails to reject the null hypothesis meaning the residuals are now homoscedastic. Due to the logarithmic transformation, the regression model's adjusted R-squared value has now increased to 78.72% as shown in Figure 7.

Dependent Variable: HOMELESS				
Method: Least Squares				
Date: 03/30/20 Time: 13:59				
Sample: 1 50				
Included observations: 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-37.86771	36.12733	-1.048174	0.3020
GINI	-33.56290	71.84803	-0.467137	0.6434
MINWAGE*COL	2.805234	0.462332	6.067578	0.0000
RENTCOST/INCOME*12	84.66933	23.02045	3.678004	0.0008
RENTVAC	0.291425	0.316509	0.920747	0.3637
POVERTY	-0.142186	0.514250	-0.276492	0.7838
UNEMP	-1.854140	1.213080	-1.528457	0.1357
VETERAN	1.229584	0.833327	1.475512	0.1493
EDUCATION	0.528255	0.270677	1.951608	0.0593
INSURANCE	0.494146	0.337906	1.462378	0.1528
DISABILITY	-1.021617	0.678236	-1.506285	0.1412
MENTAL	308.5230	227.9243	1.353621	0.1848
ALCOHOL	-159.8845	128.5921	-1.243346	0.2222
COCAINE	97.96808	275.7786	0.355242	0.7246
HEROINE	-1040.620	614.1850	-1.694310	0.0993
TEMP	-0.372303	0.102432	-3.634636	0.0009
R-squared	0.858488	Mean dependent var		14.28168
Adjusted R-squared	0.796057	S.D. dependent var		9.764168
S.E. of regression	4.409504	Akaike info criterion		6.059739
Sum squared resid	661.0868	Schwarz criterion		6.671587
Log likelihood	-135.4935	Hannan-Quinn criter.		6.292734
F-statistic	13.75087	Durbin-Watson stat		1.879925
Prob(F-statistic)	0.000000			

Figure 3

Dependent Variable: HOMELESS				
Method: Stepwise Regression				
Date: 03/30/20 Time: 14:03				
Sample: 1 50				
Included observations: 50				
Number of always included regressors: 1				
Number of search regressors: 15				
Selection method: Stepwise backwards				
Stopping criterion: p-value forwards/backwards = 0.05/0.05				
Variable	Coefficient	Std. Error	t-Statistic	Prob.*
C	-28.00565	5.131042	-5.458083	0.0000
MINWAGE*COL	2.395925	0.340897	7.028301	0.0000
RENTCOST/INCOME*12	66.70220	15.91981	4.189886	0.0001
INSURANCE	0.619686	0.268264	2.309982	0.0256
TEMP	-0.223557	0.077930	-2.868681	0.0063
UNEMP	-2.088674	0.763904	-2.734209	0.0090
R-squared	0.802335	Mean dependent var		14.28168
Adjusted R-squared	0.779873	S.D. dependent var		9.764168
S.E. of regression	4.581126	Akaike info criterion		5.993933
Sum squared resid	923.4155	Schwarz criterion		6.223376
Log likelihood	-143.8483	Hannan-Quinn criter.		6.081307
F-statistic	35.71968	Durbin-Watson stat		1.835879
Prob(F-statistic)	0.000000			

Figure 4

Heteroskedasticity Test: Breusch-Pagan-Godfrey			
Null hypothesis: Homoskedasticity			
F-statistic	2.821831	Prob. F(5,44)	0.0270
Obs*R-squared	12.14022	Prob. Chi-Square(5)	0.0329
Scaled explained SS	11.24095	Prob. Chi-Square(5)	0.0468

Figure 5

Heteroskedasticity Test: Breusch-Pagan-Godfrey			
Null hypothesis: Homoskedasticity			
F-statistic	0.745048	Prob. F(5,44)	0.5940
Obs*R-squared	3.902797	Prob. Chi-Square(5)	0.5635
Scaled explained SS	3.796246	Prob. Chi-Square(5)	0.5791

Figure 6

Finally, the model's residuals must be tested in order to ensure that they follow a normal distribution. In order to test this, a Jarque-Bera test is utilized which forms a test statistic using the sample size, sample skewness coefficient, and the kurtosis coefficient. This test has a null hypothesis that the residual is normally distributed and an alternate hypothesis that the residual is not normally distributed ("Jarque-Bera Test", 2016). This test yields a Jarque-Bera value of 1.149 and a p-value of 56.29% which is insignificant, indicating that the residual is normal in distribution. Due to a normal homoscedastic residual distribution, the model's assumption of linearity has been met and since this is a cross-sectional analysis, autocorrelation is not a factor to be concerned with. Figure 7 depicts the final multiple regression model which accounts for the ordinary least square regression assumptions of relational linearity, error homoscedasticity and normality, and adjusts for multicollinearity.

Dependent Variable: LOG(HOMELESS)
Method: Least Squares
Date: 03/30/20 Time: 14:08
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.341011	0.274322	1.243108	0.2204
MINWAGE*COL	0.136592	0.018225	7.494611	0.0000
RENTCOST/INCOME*12	3.499198	0.851123	4.111273	0.0002
INSURANCE	0.046760	0.014342	3.260273	0.0022
TEMP	-0.016121	0.004166	-3.869200	0.0004
UNEMP	-0.135627	0.040841	-3.320886	0.0018

R-squared	0.808915	Mean dependent var	2.498052
Adjusted R-squared	0.787201	S.D. dependent var	0.530935
S.E. of regression	0.244921	Akaike info criterion	0.136407
Sum squared resid	2.639404	Schwarz criterion	0.365850
Log likelihood	2.589823	Hannan-Quinn criter.	0.223780
F-statistic	37.25276	Durbin-Watson stat	2.217031
Prob(F-statistic)	0.000000		

Figure 7

To interpret the results of the final model, it is vital to remember that the response variable, Homeless, had a natural log transformation. When interpreting the results of a log-transformed dependent variable, the independent variable's coefficient must first be exponentiated before subtracting 1 and multiplying by 100. This resulting number is the percent change in the dependent variable given a one-unit change in any given independent variable (Ford, 2018).

The first parameter in this model is the y-intercept which in this case is denoted as "C", has a value of .3410 but is insignificant with a p-value of .2204. The second parameter, MinWage*CoL, is highly significant and has a coefficient of .1366. To understand what its effect on the homelessness rate is, the log-transformation should first be adjusted as described above. The calculations proceed as follows: $(e^{0.136592} - 1) * 100 = 14.64$. This means that when the MinWage*CoL figure increases by one dollar, the number of homeless per 10,000 individuals increases by 14.64%. In order to see what the percent increase in homelessness is, 14.64% must be divided by 100 in order to yield the change that happens in the number of homeless per 100 individuals. This yields 0.1464 meaning that this model predicts a state's homeless rate will increase by 0.1464% for every dollar that the minimum wage increases. While the minimum wage is often intended and expected to support the incomes of the poor, minimum wage is also thought to decrease the quantity demanded of labor resulting in unemployment (MaCurdy, 2015, p. 497). This could be one reason or explanation for why this model finds that raising the minimum wage also increases homelessness.

The third parameter, RentCost/Income*12, has a coefficient of 3.4992 which when adjusted yields: $(e^{3.499198} - 1) * 100 = 3208.89$. This result indicates that a one unit increase in RentCost/Income*12, will increase the homelessness rate per 10,000 people by 3208.89%. When divided by 100, a one unit increase in the percent of income spent on rent will increase the number of homeless per 100 individuals by 32.08%. From these calculations, it is clear that this variable is extremely weighty, especially when considering its magnitude compared to the other significant independent variables. Given its significance and large magnitude, this variable indicates that homelessness is closely tied to economic factors related to income and housing cost.

The fourth parameter, Insurance, has a coefficient of 0.0468 which when adjusted yields: $(e^{0.046760} - 1) * 100 = 4.79$. This means that when the percent of civilian noninstitutionalized public without health care increase one unit, the homelessness rate per 10,000 individuals increases by 4.79% or in other terms, the homelessness rate per 100 individuals increases by .0479%. Whether this result implies that possession of health insurance is a key factor in homelessness rates due to the insurance's direct health benefits or due to a proxy relationship is ambiguous; it is clear, however, that there is a significant relationship with a moderate impact.

The fifth parameter, Temp, which represented the average winter temperature, had a coefficient of -0.0161 which when adjusted yields: $(e^{-0.016121} - 1) * 100 = -1.60$. This means that an increase of one degree in a state's average winter temperature yields a 1.6% decrease in their homelessness rate per 10,000 individuals, or a 0.016% decrease in the homelessness rate per 100 individuals. As the dependent variable is a point-in-time homelessness count which is conducted in January, I expected the temperature to have a positive coefficient as I thought the unsheltered homeless would prefer warmer locations to colder locations. The results of this analysis did not support this hypothesis; however, the magnitude and/or significance of this variable might differ in a regression that analyzes only the unsheltered homeless rather than the overall homeless population.

The sixth and final parameter, Unemp, which represents the state unemployment rate of the civilian labor force, has a coefficient of -0.1356. The log-adjustment yields: $(e^{-0.135627} - 1) * 100 = -12.68$. This result can be interpreted to mean that a one percent increase in the unemployment rate yields a 12.68% decrease in the homeless rate per 10,000 people or a .1268% decrease in the homeless rate per 100 people. This variable also yielded a value which I did not expect as unemployment is often considered to be a cause of homelessness.

Conclusion

The purpose of this study was to take state-level data and find explanatory variables which could help researchers predict the rate of homelessness in any given state. Out of the 16 independent variables that the original model started with, only five were significant: the minimum wage rate adjusted for cost of living, the percent of income used on housing, the percent of population without health insurance, the average winter temperature, and the unemployment rate. These highly significant variables became the core of my model and were able to predict the homelessness rates relatively well due to a high R-squared value.

As far as policy implications go, all the variables in this model can be altered through legislation and public/private initiatives except for temperature. When analyzing these variables and their meanings from a public policy lens, there are many ideas and suggestions that policy makers could derive from this study, but an important element of this research is to remember that significant variables are not necessarily causal. As such, policy could address some of the factors, however, further research on these variables is needed to better understand their dynamics.

The most compelling component of this study, in terms of independent variables, was the variable which dealt with the percent of a household's income spent on rent. This variable had a coefficient with an extremely large magnitude implying that a relatively small change in the percent of income spent on housing could yield great results in decreasing the homelessness rate. Policies which could address this range from federal rental assistance to the creation of public housing to income subsidies. While these approaches may not address all the core needs of the homeless, according to this research, it could have the largest impact on homelessness rates compared to the other significant variables. In order to truly understand this dynamic, however, whole studies could and should be conducted solely on the impact of government support on homeless populations.

Given the opportunity to continue this research into the future, I would make a few alterations to my research design, however, I would maintain my original scope. The purpose of this research was to find factors which predict homelessness at a state-wide level and while looking at homelessness rates at a county or metropolitan level may yield different results with intriguing conclusions, I am still interested in the purpose of the original research. As such, the scope would stay the same, but I would try to segment my homeless populations more. Rather than regressing all of these factors against the overall homeless population, I believe it would be beneficial to regress the independent variables against the chronically homeless in comparison to the temporarily homeless populations as well as against the sheltered homeless and the unsheltered homeless populations. By dividing the populations up, trends that are specific to one group but not the other could emerge. This method might clear up the confusion produced by the sign of the magnitude on the unemployment and temperature variables.

While this research produced more questions that need to be answered in order for society to properly and efficiently respond to the homelessness crisis that is present in the United States, it was able to produce a significant model which can assist researchers in predicting state homelessness rates. Going forward, more research is necessary to clarify elements of this model as well as introduce new concepts to the research design. Homelessness is more than just a choice; it is a condition brought about by economic and environmental factors. By furthering research such as this, society can be more understanding and take more effective steps in curbing homelessness and addressing the needs of this vulnerable population.

References

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics*, *05*(07), 754–767. doi: 10.4236/ojs.2015.57075

Astivia, O. L. O., & Zumbo, B. D. (2019). Heteroskedasticity in Multiple Regression Analysis: What it is, How to Detect it and How to Solve it with Applications in R and SPSS. *Practical Assessment, Research, and Evaluation*, *24*(1), 1–16. doi: https://doi.org/10.7275/q5xr-fr95

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287–1294. doi: 10.2307/1911963

Bush, H., & Shinn, M. (2017). Families' Experiences of Doubling Up After Homelessness. *Cityscape (Washington, D.C.)*, *19*(3), 331–356.

Ford, C. (2018, August 17). Interpreting Log Transformations in a Linear Model. Retrieved March 29, 2020, from https://data.library.virginia.edu/interpreting-log-transformations-in-a-linear-model/

Committee on an Evaluation of Permanent Supportive Housing Programs for Homeless Individuals, Science and Technology for Sustainability Program, Policy and Global Affairs, Board on Population Health and Public Health Practice, Health and Medicine Division, & National Academies of Sciences, Engineering, and Medicine. (2018). *Permanent supportive housing: evaluating the evidence for improving health outcomes among people experiencing chronic homelessness* [National Center for Biotechnology Information]. Washington, DC: The National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/

Council of Economic Advisers. The State of Homelessness in America, The State of Homelessness in America (2019). Retrieved from https://www.whitehouse.gov/wp-content/uploads/2019/09/The-State-of-Homelessness-in-America.pdf

Eggers, F. J., & Moumen, F. Analysis of Trends in Household Composition Using American Housing Survey Data, Analysis of Trends in Household Composition Using American Housing Survey Data (2013). Retrieved from https://www.huduser.gov/portal//publications/pdf/AHS_HouseholdComposition_v2.pdf

Evans, W. N., Phillips, D. C., & Ruffini, K. (2019). *Reducing and Preventing Homelessness: A Review of the Evidence and Charting a Research Agenda*. Abdul Latif Jameel Poverty Action Lab. Retrieved from https://www.povertyactionlab.org/sites/default/files/publications/reducing-preventing-homelessness-review-research-agenda.pdf

Foscarinis, M. (1996). Downward Spiral: Homelessness and Its Criminalization. *Yale Law & Policy Review, 14*(1), 1-63. Retrieved March 16, 2020, from www.jstor.org/stable/40239449

Fowler, P. J., Hovmand, P. S., Marcal, K. E., & Das, S. (2019). Solving Homelessness from a Complex Systems Perspective: Insights for Prevention Responses. *Annual Review of Public Health*, *40*(1), 465–486. doi: 10.1146/annurev-publhealth-040617-013553

Garrett D. G. (2012). The business case for ending homelessness: having a home improves health, reduces healthcare utilization and costs. *American health & drug benefits*, *5*(1), 17–19.

Government Accountability Office. Homelessness: A Common Vocabulary Could Help Agencies Collaborate and Collect More Consistent Data, Homelessness: A Common

Vocabulary Could Help Agencies Collaborate and Collect More Consistent Data (2010). Retrieved from https://www.gao.gov/new.items/d10702.pdf
Henry, M., Mahathey, A., Morrill, T., Robinson, A., Shivji, A., Watt, R., & Abt Associates. 2018 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S., 2018 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S. (2018). Retrieved from https://files.hudexchange.info/resources/documents/2018-AHAR-Part-1.pdf
Hopper, K., Shinn, M., Laska, E., Meisner, M., & Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American journal of public health*, *98*(8), 1438–1442. https://doi.org/10.2105/AJPH.2005.083600
IHS Global Inc. (2019, April 4). Stepwise Least Squares Regression. Retrieved March 28, 2020, from http://www.eviews.com/help/helpintro.html#page/content/Regress2-Stepwise_Least_Squares_Regression.html
Jarque-Bera Test. (2016, May 7). Retrieved March 29, 2020, from https://www.statisticshowto.datasciencecentral.com/jarque-bera-test/
Johnsen, S., & Watts, B. (2014). *Homelessness and Poverty: Reviewing the Links*. Institute for Social Policy, Housing, Environment and Real Estate. Retrieved from https://pureapps2.hw.ac.uk/ws/portalfiles/portal/7467281/Homelessness_Poverty_FullReport.pdf
Joint Center for Housing Studies of Harvard University. (2015). *America's Rental Housing Expanding Options for Diverse and Growing Demand*. Harvard Kennedy School. Retrieved from https://www.jchs.harvard.edu/sites/default/files/America's Rental Housing 2015_WEB.pdf
Macurdy, T. (2015). How Effective Is the Minimum Wage at Supporting the Poor? *Journal of Political Economy*, *123*(2), 497–545. doi: 10.1086/679626
McKinney-Vento Homeless Assistance Act of 1987, 42 U.S.C. § 11301.
Metraux, S., Fargo, J. D., Eng, N., & Culhane, D. P. (2018). Employment and Earnings Trajectories During Two Decades Among Adults in New York City Homeless Shelters. *Cityscape: A Journal of Policy Development and Research*, *20*(2), 171–2020.
Missouri Economic Research and Information Center. (2020). Cost of Living Data Series. Retrieved March 19, 2020, from https://meric.mo.gov/data/cost-living-data-series
National Alliance to End Homelessness. (2015). *Ending Chronic Homelessness Saves Taxpayers Money*. endhomelessness.org. Retrieved from http://endhomelessness.org/wp-content/uploads/2017/06/Cost-Savings-from-PSH.pdf
Nooe, R. M., & Patterson, D. A. (2010). The Ecology of Homelessness. *Journal of Human Behavior in the Social Environment*, *20*(2), 105–152.
Office of Community Planning and Development. 2017 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S., 2017 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S. (2017). Retrieved from https://files.hudexchange.info/resources/documents/2017-AHAR-Part-1.pdf
Office of Community Planning and Development. 2017 AHAR: Part 2 - Estimates of Homelessness in the U.S., 2017 AHAR: Part 2 - Estimates of Homelessness in the U.S. (2018). Retrieved from https://files.hudexchange.info/resources/documents/2017-AHAR-Part-2.pdf

Osborn, L. (n.d.). Winter Temperature Averages for Every State. Retrieved March 20, 2020, from https://www.currentresults.com/Weather/US/average-state-temperatures-in-winter.php

Parsell, C., & Parsell, M. (2012). Homelessness as a Choice. *Housing, Theory, and Society*, *29*(4), 420–434. doi: 10.1080/14036096.2012.667834

Quigley, J. M., & Raphael, S. (2001). The Economics Of Homelessness: The Evidence From North America. *European Journal of Housing Policy*, *1*(3), 323–336. doi: 10.1080/14616710110091525

Shinn, M. (2010). Homelessness, Poverty and Social Exclusion in the United States and Europe. *European Journal of Homelessness*, *4*, 19–44. Retrieved from https://www.feantsaresearch.org/download/article-1-23498373030877943020.pdf

Smith, G. (2018). Step away from stepwise. *Journal of Big Data*, *5*(1). doi: 10.1186/s40537-018-0143-6

Substance Abuse and Mental Health Services Administration (2018). *2014-2016 Substate Estimates of Substance Use and Mental Illness* [Data Set]. Retrieved from http://www.samhsa.gov/data/report/2014-2016-nsduh-substate-region-estimates-excel-tables-and-csv-files

U.S. Bureau of Labor Statistics. (2020, March 4). Unemployment Rates for States, Annual Averages. Retrieved March 19, 2020, from https://www.bls.gov/lau/lastrk17.htm

U.S. Census Bureau (2019a), Rental Vacancy Rate for the United States [RRVRUSQ156N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/RRVRUSQ156N, March 13, 2020.

U.S. Census Bureau (2019b). *Table 1. Rental Vacancy Rates by State: 2005-present* [Data Set]. Retrieved from https://www.census.gov/housing/hvs/data/rates.html

U.S. Census Bureau (2018a). *2017 American Community Survey 1-Year Estimates (S0102)* [Data Set]. Retrieved from https://www.census.gov/newsroom/press-kits/2018/acs-1year.html

U.S. Census Bureau (2018b). *2017 American Community Survey 1-Year Estimates (B19083)* [Data Set]. Retrieved from https://www.census.gov/newsroom/press-kits/2018/acs-1year.html

U.S. Census Bureau (2018c). NST-EST2018-01: *Table 1. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2018* [Data Set]. Retrieved from https://www.census.gov/newsroom/press-kits/2018/pop-estimates-national-state.html

U.S. Census Bureau (2017). *Percentage of People in Poverty by State Using 2- and 3-Year Averages: 2014-2015 and 2016-2017* [Data Set]. Retrieved from https://www.census.gov/library/publications/2018/demo/p60-263.html

U.S. Department of Veteran Affairs (2019). *Table 8L: VETPOP2016 Living Veterans by State, Race/Ethnicity, Gender, 2015-2045* [Data Set]. Retrieved from https://www.va.gov/vetdata/veteran_population.asp

U.S. Department for Housing and Urban Development. (2020). AHAR Reports. Retrieved from https://www.hudexchange.info/homelessness-assistance/ahar/

U.S. Department for Housing and Urban Development (2019). 2007 - 2019 HIC Data by State [Data set]. United States Department of Housing and Urban Development. Retrieved from https://www.hudexchange.info/resource/3031/pit-and-hic-data-since-2007/

U.S. Department for Housing and Urban Development. Point-in-Time Count Methodology Guide, Point-in-Time Count Methodology Guide (2014). Retrieved from https://files.hudexchange.info/resources/documents/PIT-Count-Methodology-Guide.pdf

U.S. Department of Labor. (2020, January). Changes in Basic Minimum Wages in Non-Farm Employment Under State Law: Selected Years 1968 to 2019. Retrieved March 19, 2020, from https://www.dol.gov/agencies/whd/state/minimum-wage/history

U.S. Interagency Council on Homelessness. Key Federal Terms and Definitions of Homelessness Among Youth, Key Federal Terms and Definitions of Homelessness Among Youth (2018). Retrieved from https://www.usich.gov/resources/uploads/asset_library/Federal-Definitions-of-Youth-Homelessness.pdf

U.S. Interagency Council on Homelessness. Ending Chronic Homelessness in 2017, Ending Chronic Homelessness in 2017 (2015). Retrieved from http://dev2.usich.gov/news/ending-chronic-homelessness-in-2017

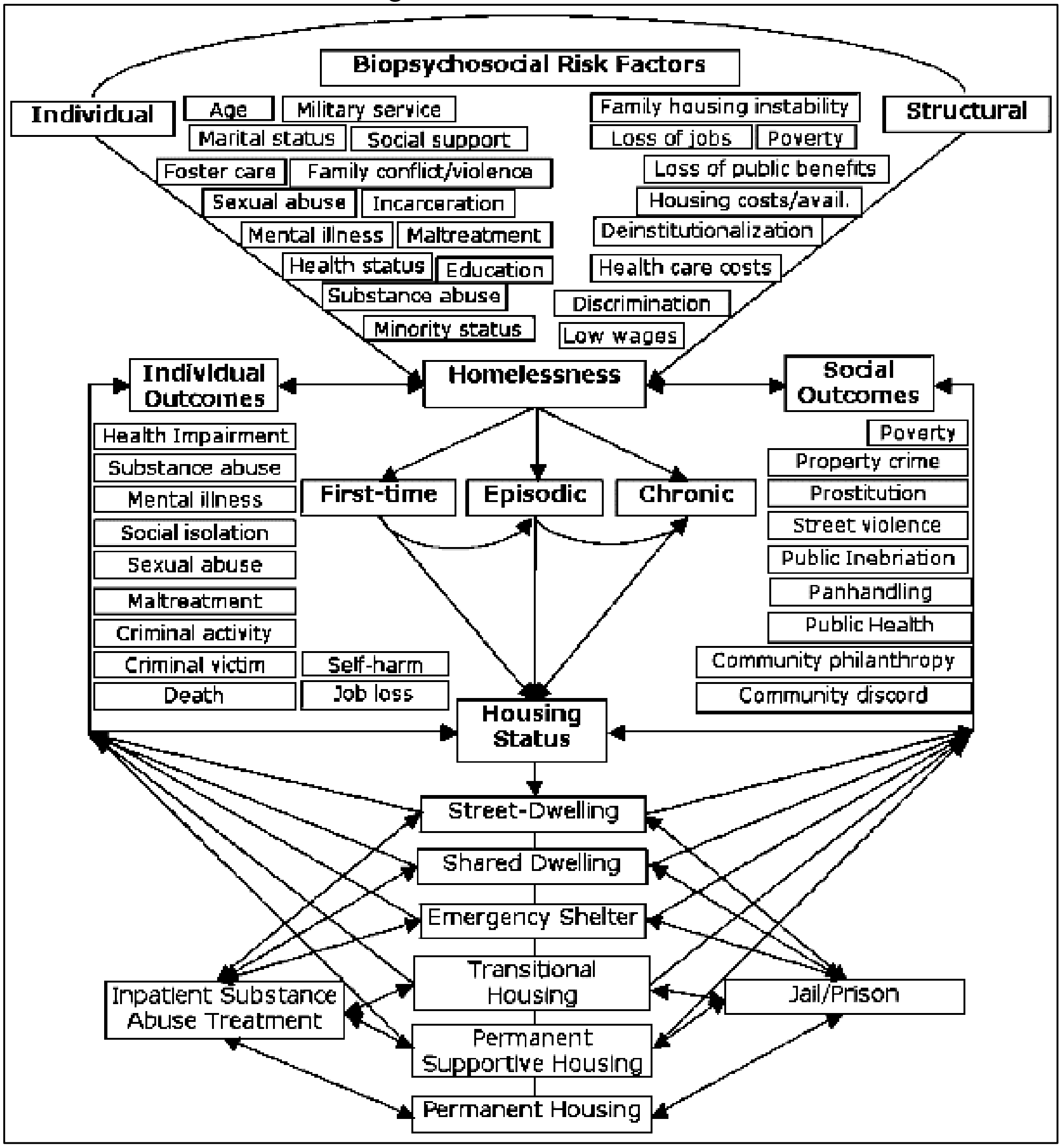
U.S. Interagency Council on Homelessness. Investing in the End of Homelessness: The President's 2017 Budget, Investing in the End of Homelessness: The President's 2017 Budget (2016). Retrieved from https://www.usich.gov/resources/uploads/asset_library/2017_Budget_USICH_Homelessness_Fact_Sheet_final.pdf

Watt, R., Henry, M., Rosenthal, L., Shivji, A., & Abt Associates, undefined. The 2017 Annual Homeless Assessment Report (AHAR) to Congress, The 2017 Annual Homeless Assessment Report (AHAR) to Congress (2017). Retrieved from https://files.hudexchange.info/resources/documents/2017-AHAR-Part-1.pdf

Wright, J. D., & Rubin, B. A. (1991). Is homelessness a housing problem? *Housing Policy Debate*, *2*(3), 937–956. doi: 10.1080/10511482.1991.9521078

Appendix A

Ecological Model of Homelessness



Ecological model of homelessness. Reprinted from "The Ecology of Homelessness," by R. M. Nooe & D. A. Patterson, 2010, Journal of Human Behavior in the Social Environment, 20(2), 107.

Appendix B
Variable Names, Descriptions, and Sources

Dependent Variable Name	Variable Description	Source/Citation
Homeless	2017 state homelessness rate (measured in homeless individuals per 10,000 individuals); found by dividing each state's 2017 point-in-time homeless count by its respective population.	HUD (2019) & U.S. Census Bureau (2018c)
Explanatory Variable Name	**Variable Description**	**Source/Citation**
CoL	2019 composite cost of living index by state. These values were divided by 100 in order to provide a decimal.	Missouri Economic Research and Information Center (2020)
Gini	2017 annual index of each state's income inequality on a scale of 0-1 where 0 represents perfect equality and 1 equals perfect inequality	U.S. Census Bureau (2018b)
MinWage	2017 active minimum wage across all states for non-farm employment.	U.S. Department of Labor (2020)
Income	2016 annual lowest quintile of household income	2016 American Community Survey (Prosperity Now, n.d.)
RentCost	2017 median gross monthly rental cost	U.S. Census Bureau (2018a)
RentVac	2017 average percent of rental units in each state which are vacant and available to rent	U.S. Census Bureau (2019b)
Poverty	2016/2017 average poverty rate of state population	U.S. Census Bureau (2017)
Unemp	2017 average unemployment rate of state civilian labor force	U.S. Bureau of Labor Statistics (2020)
Veteran	2017 annual percent of living veterans to state population over 18	U.S. Department of Veteran Affairs (2019)
Education	2017 annual percent of each state's population over the age of 25 with an education attainment no greater than high school completion	U.S. Census Bureau (2018a)
Insurance	2017 annual percent of civilian noninstitutionalized public without health care	U.S. Census Bureau (2018a)
Disability	2017 annual percent of civilian noninstitutionalized public between the ages of 18 and 64 with a disability	U.S. Census Bureau (2018a)
Mental	2014-2016 average annual percent of individuals 18 and over which had a serious, diagnosable mental illness	Substance Abuse and Mental Health Services Administration [SAMHSA] (2018)
Alcohol	2014-2016 average annual percent of individuals 12 and older which had an alcohol dependence/abuse disorder in the past year	SAMHSA (2018)
Cocaine	2014-2016 average annual percent of individuals 12 and older which used cocaine in the past year	SAMHSA (2018)
Heroine	2014-2016 average annual percent of individuals 12 and older which used heroine in the past year	SAMHSA (2018)
Temp	2017 average winter temperature (°F) by state	Current Results (Osborn, n.d.)

Appendix C
Original Correlation Matrix

	Gini	*MinWage*CoL*	*Income*CoL*	*RentCost*CoL*	*RentVac*	*Poverty*	*Unemp*	*Veteran*	*Education*	*Insurance*	*Disability*	*Mental*	*Alcohol*	*Cocaine*	*Heroine*	*Temp*
Gini	1															
MinWage*CoL	0.039346	1														
Income*CoL	-0.23148	0.827201289	1													
RentCost*CoL	0.029353	0.877837686	0.930302739	1												
RentVac	0.003943	-0.520515168	-0.343590457	-0.36456586	1											
Poverty	0.466219	-0.343523309	-0.589054142	-0.386774638	0.415427	1										
Unemp	0.514255	0.010264707	-0.14779399	0.026831458	0.171839	0.660542	1									
Veteran	-0.40842	-0.163921719	-0.098563911	-0.166128173	0.211841	0.139423	-0.01041	1								
Education	0.453054	-0.399316072	-0.547667899	-0.409480862	0.348624	0.729495	0.529607	0.083939	1							
Insurance	-0.01613	-0.476320718	-0.361090186	-0.305104912	0.436835	0.408568	0.29313	0.284615	0.177332	1						
Disability	0.158235	-0.350793339	-0.592266389	-0.49682673	0.308859	0.649966	0.38995	0.435778	0.712361	0.145866	1					
Mental	-0.29895	-0.246884657	-0.36745046	-0.425833385	-0.07873	0.064757	-0.12818	0.334704	0.160079	-0.15613	0.554213	1				
Alcohol	-0.16596	0.280003072	0.220601789	0.131853135	-0.18239	-0.25487	-0.19292	0.076089	-0.34165	-0.22594	-0.29566	-0.06802	1			
Cocaine	0.226814	0.649927845	0.485371776	0.5483224	-0.57186	-0.26221	0.118786	-0.1321	-0.30461	-0.3647	-0.2276	0.025109	0.464623	1		
Heroine	-0.00841	0.344298954	0.363982538	0.241878069	-0.42507	-0.31233	0.167587	0.059575	-0.05129	-0.32064	-0.0542	0.128104	0.16022	0.50012	1	
Temp	0.485074	0.013902242	-0.035044212	0.232994305	0.162511	0.436362	0.361812	-0.03281	0.374079	0.265124	0.170186	-0.28023	-0.46265	-0.06098	-0.3185	1