Do People Living in Rural Areas Have Less Severe Depression Problems? Evidence from NHIS 2019 Survey Data

Gaonan You

Department of Economics, Dalhousie University, Halifax, Canada

*Corresponding author: Gaonan.You@dal.ca

Abstract. Although many papers confirm that people living in rural areas are less likely to suffer from depression than those living in urban areas, most of them employed a straightforward linear regression, which is not convincing. Using data from the 2019 National Health Interview Survey (NHIS). This paper first identifies a proxy for depression level through Pearson's Chi-Squared test. Then, using OLS, this paper determined that place of residence had an impact on depression prevalence. To reduce the bias of the estimated the effect of place, this paper employs a propensity score matching method. Finally, the matched sample was retested to see if residence increased the risk of depression. This study suggests that there is a correlation between where you live and depression.

Keywords: depression level; Urban-rural; mental health; America.

1. Introduction

Mental health refers to individual's emotional, psychological, and social well-being. It is as important as physical health to individual's overall health. European countries have seen an increase in depression, which has become a very common condition [1]. Rates in Canada are lower than in the United States, but on par with those in European countries [2]. However, the situation in America is the worst. According to the Centers for Disease Control and Prevention [3], mental health has become one of the most common illnesses in the United States. More than 50 percent of Americans have been diagnosed with a mental illness or disorder at some point in their lives. The mental health is interrelated with physical health, and it also has impact on the economy as the worse mental health can lead to the loss of productivity and increased expenditure in medicare. Therefore, understanding mental health and which factors influence mental health are of great importance.

The economic impact of mental illness includes lost productivity and increased use of treatment resources. If mental and physical illness coexist, the longer the disease lasts, the more economics cost to society will spend. According to Colorado, an American insurance company, people with mental illness spends more than doubled of health care service than healthy people. This is because some patients with mental illness need treatment for other chronic diseases, others have unexplained illness, they tend to get more medical tests. Mental illness can affect the course of an individual's life and even spread to families and communities [4].

The importance of mental health is obvious, but many surveys are based only on people living in urban areas or discuss the mental health of rural and urban residents separately. For example, when examining the impact of COVID-19 on residents' well-being, the report focused only on urban residents. Because of the lack of well-being of people living in the countryside, the psychological state of these people is difficult to speculate. Compared with urban areas, residents in rural areas face more difficulties. Such as high poverty, low employment rates and poor health care [5]. What is more, the COVID-19 pandemic that began in 2020 has undoubtedly exacerbated the gap of social status and income. Due to the lack of comparison of mental health of residents in the two places, the factors affecting the difference in mental health of residents in the two places are ambiguous. Do cities cause mental illness? The answer in the literatures is yes. This reflects the shortcomings of cities, which can be persistent [6]. If this discrepancy is not addressed, it will lead to insufficient information and the policies proposed by the government will be ineffective for residents in rural areas.

The biological and genetic causes of depression have been studied by many researchers and there's still much ongoing research. Nevertheless, the social characteristics are also considered as risk factors of depression and the urban or rural residence is one of them. According to Maslow's hierarchy of needs [7], people are motivated by five basic categories of needs: psychological, safety, love and belonging, esteem, and self-actualization. The five categories are often depicted as a pyramid with the categories sorting from lower to upper level. Every person has the desire to move up toward the self-actualization. The rapid development of urbanization in the United States since the 1950s provides people more resources to satisfy their needs. People living in urban areas have greater access to job market, better education, social location and entertainment resources. It is reasonable to assume that living in urban area is both culturally and economically beneficial. However, many literatures have found living in urban area could increase the risk of depression. Crowell et al. [8] use the Piedmont Health Survey data from North Carolina and find that urban residents are more vulnerable to depression and the risk of major depression1 decreases for some specific demographic subgroups living in rural area. They conclude that rural residence can be a buffer against major depression. Similar results are also found in Scotland. Instead of focusing on individual level depression data, McKenzie, Murray, and Booth [9] utilize macro-level data and use drug prescriptions within certain data zones as a proxy of depression level in an area. More urban residents are found to be associated with higher usage of prescription.

Most of the literatures studying the urban and rural differences in depression level employ a straightforward qualitative analysis or linear regression framework or logistic regression. These models can barely be explained causally. This paper uses the 2019 National Health Interview Survey (NHIS) data to revisit this question by employing a propensity score match- ing method. I find that the depression becomes severer if individuals move to rural area. The probability of taking prescription for depression also increases by about 3% if individuals move to rural area.

The next section describes the NHIS data and makes some descriptive analysis. Section 3 shows preliminary analysis for OLS regression in multiple specifications. It also introduces how this paper apply the propensity score matching method and estimate the effect of living in urban areas on the depression severity. Section 4 concludes the analysisand discusses the potential future work.

Inevitably, the results of the survey are slightly biased. The difference in the results came from the fact that people in rural areas were less likely to be surveyed if they were depressed [10].

2. Data

The NHIS data is a cross sectional household interview survey targeting at the residents in the 50 states and the District of Columbia. The whole country is divided into 1,689 geographic areas. For some states, the geographic areas are divided into two strata while only one stratum is formed for the remaining states. Clusters of addresses are defined within each of these areas. Clusters are selected from the strata and the number of clusters selected is proportional to the number of clusters in the strata.

This paper only uses the 2019 NHIS sample adult survey, which has 31,997 observations. The independent variable of interest is the urban-rural classification. This variable has four categories, and this paper only keeps those observations classified as living in large central metro or nonmetropolitan. Other variables related with social characteristics are also selected as control variables, including physical health status, family income, gender, education attainment, marital status, and age. Education attainment in the original survey has twelve categories. This paper re-encodes the twelve categories to four categories: high school or less, college equivalent, bachelor, graduate2. Marital status has three categories. This paper combines the first two categories together and the new defined variable only tell whether the respondent lives with a partner or not. Because education attainment is included, this paper further restricts the observations to those whose age are from 24 to 65 (inclusive). The dependent variable of interest is severity of depression which is measured by the eight-item Patient Health Questionnaire depression scale (PHQ-8). The scale is recoded into four categories: none/minimal, mild, moderate, and severe3. Another variable which tells whether the respondent takes prescription for depression or not is also selected as a dependent variable following McKenzie, Murray, and Booth's [7]work. This variable can be used as an indicator of the severity of depression. It is likely that individuals suffering from more severe depression tend to take prescription.

Dropping observations with null values, the final sample size is 9398. Table 1 presents the basic summary statistics for the dependent variables and continuous independent variables.

Table 1. Summary Statistics by Place of Residence

	Urban (n = 6358)		Rural (n = 3040)		
	Mean	Std. Dev	Mean	Std. Dev	Difference
Dependent Variable					
Depression	1.27	0.66	1.37	0.77	-0.10^{***}
Prescription	0.09	0.29	0.14	0.34	-0.20^{***}
Independent Variable					
Age	44.16	12.19	47.00	12.23	-2.84^{***}
Female	0.52	0.50	0.53	0.50	-0.01
Partner	0.56	0.50	0.62	0.49	-0.06^{***}
Physical Health Status: Poor	0.03	0.16	0.05	0.21	-0.02^{***}
Physical Health Status: Fair	0.10	0.30	0.14	0.35	-0.04^{***}
Physical Health Status: Good	0.24	0.43	0.30	0.46	-0.06^{***}
Physical Health Status: Very Good	0.34	0.48	0.33	0.47	0.01
Physical Health Status: Excellent	0.29	0.45	0.18	0.38	0.11^{***}
Income: Less than $34,999	0.23	0.42	0.33	0.47	-0.10^{***}
Income: $35,000-$49,999	0.12	0.32	0.14	0.35	-0.02^{***}
Income: $50,000-$74,999	0.18	0.38	0.20	0.40	-0.02^{***}
Income: $75,000-$99,999	0.13	0.33	0.14	0.34	-0.01
Income: More than $100,000	0.34	0.48	0.19	0.39	0.15^{***}
Education: High school or less	0.27	0.44	0.44	0.50	-0.17^{***}
Education: College equivalent	0.25	0.44	0.33	0.47	-0.08^{***}
Education: Bachelor	0.29	0.45	0.15	0.35	0.14^{***}
Education: Graduate or higher	0.18	0.39	0.08	0.27	0.10^{***}
Note:					*p<0.1; **p<0.05; ***p<0.01

3. Empirical results

To study the effect of place of residence on the severity of depression, this paper first uses the linear regression

$$Y_i = \alpha + \delta rural_i + \beta X_i + e_i \tag{1}$$

Where *Yi* is dependent variables which are the PHQ-8 scale depression level of individual i, rurali is a dummy variable which equals one if the individual lives in rural area, *Xi* is a vector of control variables including age, age square, gender, marital status, physical health status, income class, and level of education. The parameter δ captures the effect of place of residence on the depression level.

Another variable, whether the individual takes prescription for depression, is also used as a dependent variable for the regression. The prescription can be used as a proxy for depression level. Intuitively, individuals having severer depression problems are more likely to take prescription. Table 2 summarizes the proportions of individuals taking prescription for depression at different depression levels. The proportion is increasing with the depression level. The Pearson's Chi-squared test rejects the null hypothesis of equal proportion for the four groups at 5% level. Therefore, I conclude that the use of prescription can be used as a proxy for depression level.

Table 2. Pearson's Chi-squared Test of Equal Proportions

	PHQ-8 Scale Depression Level				
	None/Minimal	Mild	Moderate	Severe	χ^2
Proportions of individuals taking prescription	0.0498	0.2482	0.4262	0.5364	1532.1

Table 3 summarizes the regression results. Heteroskedasticity-robust standard errors are reported in the parentheses. Without including control variables, the place of residence has a statistically significant effect on the depression level at 5% level. Individuals living in rural area are more likely to have severer depression problems. But the effect becomes insignificant at 5% level if control variables are included. The magnitudes of the coefficient estimates from both regressions are quite small. It seems that the effect of place of residence on depression level is economically insignificant. However, the interpretation of the coefficient estimate is somewhat inappropriate. The linear regression model assumes that the dependent variable, PHQ-8 scale depression level, is continuous. But the numbers in the dataset only stands for different categories so they are not cardinal numbers. Instead, they are ordinal numbers. The differences between the categories are unknown and it is unreasonable to assume they are equal, so neither the estimate 0.096 nor 0.028 tells us the actual effect of place of residence on the depression severity. Nevertheless, the regression results still provide some insights into the difference of depression levels between individuals in rural and urban area. For regressions in the third and fourth column in Table 3, the dependent variable prescription is a dummy variable. Therefore, the linear regression model is in fact the linear probability model. Without including control variables, the coefficient estimate is statistically significant at 5% level. It indicates that the probability of taking prescription for depression increases 4.7% when individuals moving from urban area to rural area. Although there's a small decrease in magnitude after including control variables, the coefficient estimate is still statistically significant at 5% level and indicates a 3% probability increase of taking prescription.

Although the estimates indicate that living in rural area has a negative impact on in- dividuals' depression level, the model can hardly be interpreted causally. The problem of self-selection may weaken the result concluded from OLS estimates. In other words, unob- served factors, such as increase in crime rates and environmental degradation, may cause individuals move to rural communities and these factors may also contribute to the worse Mental health status. In the past, people moved to cities to look for job opportunities and better resources for education and so on. But the concentration of huge population in cities also brought the problem of scarcity of resources. People face with increasing living cost and crime rates. The environment in urban area is also deteriorating. The past years have witnessed redistribution of population not only in the United States but also many other countries. The development of technology in internet and communication also makes it pos- sible to work from home. People begin to reconsider the costs and gains of living in cities and many of them have chosen to move to rural area. This trend becomes even clearer since the outbreak of COVID-19.

Table 3: OLS: Estimated Effects of living in Rural Area

	Dependent variable:			
	depression		prescription	
	(1)	(2)	(3)	(4)
Rural	0.096***	0.028*	0.047***	0.030***
	(0.016)	(0.015)	(0.007)	(0.007)
Constant	1.274***	2.315***	0.091***	0.195***
	(0.008)	(0.120)	(0.004)	(0.052)
Controls	No	Yes	No	Yes
Observations	9,398	9,398	9,398	9,398
Adjusted R^2	0.004	0.190	0.005	0.082

Note: *p<0.1; **p<0.05; ***p<0.01

To address the potential issues, this paper employs a propensity score matching method to reduce the bias of the estimated effect of place of residence on the depression level. The propensity score matching method is typically used when the treated units and untreated units are not comparable or the selection of comparable units is difficult. Table 1 shows that the mean differences of most variables are significant and this indicates that the covariates are not balanced. The goal of propensity score matching is to produce covariate balance so that the distributions of covariates in the treated and untreated group are approximately equal. Then the difference in means can be used to estimate the treatment effect. Ho et al. [7] introduce a propensity score matching method which can reduce model dependence and produce more accurate estimated causal effects. This paper will employ their method to preprocess the observational data and then apply the linear regression model again with the matched sample.

A probit model

$$\Pr(rural_i = 1|X_i) = \Phi(\alpha + \beta X_i) \tag{2}$$

Is used to estimate the propensity score. $\Phi(\cdot)$ is the standard normal cumulative distribution function and Xi is the vector of control variables. Propensity scores are the fitted values from the probit regression and they are used to calculate the conditional probability of living in rural area regardless of actual place of residence. Figure 1 shows the distributions of propensity scores for the two groups. We can visually check the common support condition which is required to calculate the treatment effect. It's apparent that the distributions overlap between the two groups and no units are outside the range of the common support. Therefore, it's ensured that there will be adequate matches between treated and untreated units.

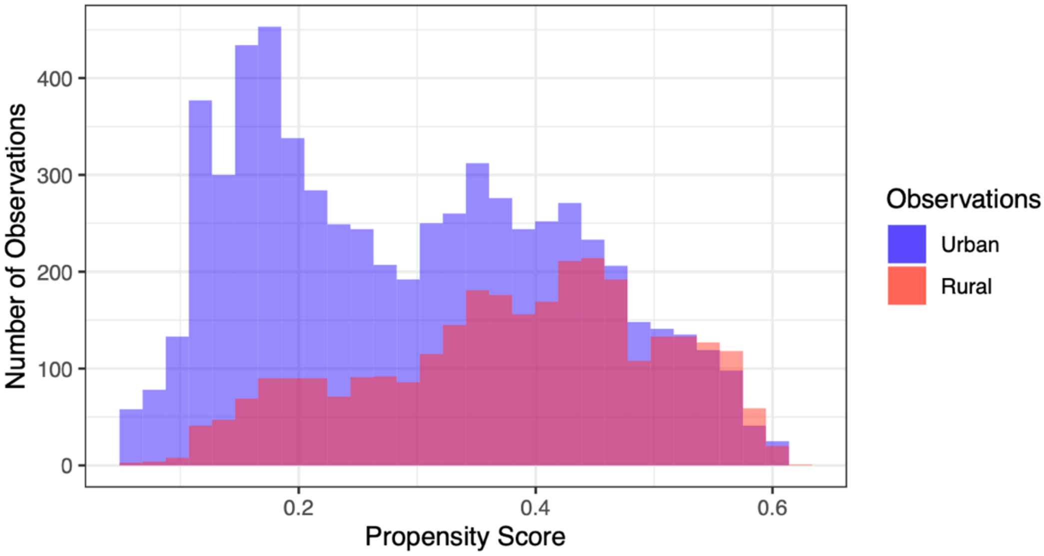

Figure 1. Histogram of Propensity Score

Next, a full matching algorithm is used so that every treated unit is matched to at least one untreated unit and every untreated unit is matched to at least one treated unit. Figure 2 visualizes the distribution of propensity score of the matched sample. Because full matching is used, there's no unmatched treated or untreated unit.

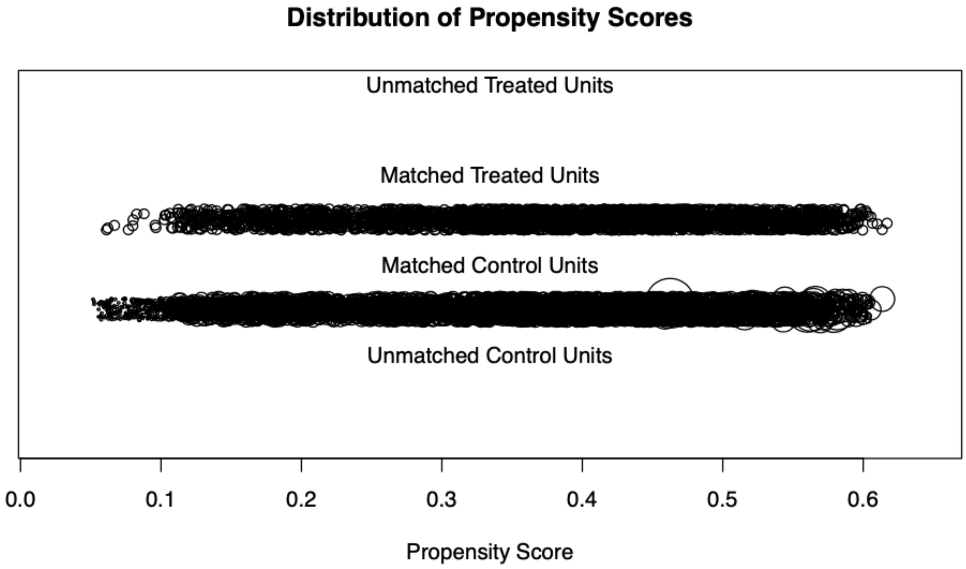

Figure 2. Distribution of Propensity Scores for the Matched Sample

Table 4 summarizes the standard mean differences of characteristics between the un- treated and treated group. A value of standard mean difference which is close to zero indicates good balance of the characteristic. Comparing the standard mean differences of characteristics before and after matching, it's apparent that the balance is far better after matching as the values are all closer to zero. Figure 3 visualizes the covariates balance before and after matching. The dotted line in the plot represents the threshold 0.1. It's easy to see that the balance is quite poor before matching. The full matching improves the balance on all characteristics as almost all of them are within the threshold.

Table 4. Standard Mean Difference of Characteristics Before and After Matching

	Standard Mean Difference	
	Unmatched	Matched
Distance	0.6834	0.0002
Age	0.2323	-0.0219
Age square	0.2311	-0.0215
Female	0.0192	-0.0060
Partner	0.1294	0.0341
Income		
Less than $34,999	0.1958	-0.0139
$35,000-$49,999	0.0744	0.0124
$50,000-$74,999	0.0647	-0.0092
$75,000-$99,999	0.0266	-0.0073
More than $100,000	-0.3892	0.0213
Education		
High school or less	0.3493	0.0176
College equivalent	0.1559	-0.0186
Bachelor	-0.4099	-0.0092
Graduate	-0.3716	0.0118
Health		
Poor	0.0866	0.0045
Fair	0.1357	-0.0273
Good	0.1254	0.0017
Very good	-0.0337	0.0132
Excellent	-0.2794	0.0043

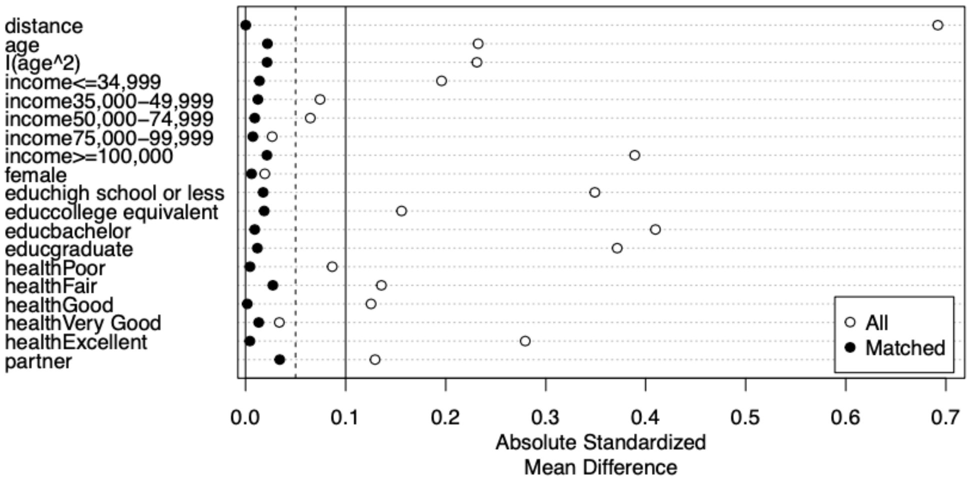

Figure 3. Love Plot of Covariate Balance

Now this paper use the linear regression model again to estimate the effect of living in rural area on depression but with the matched sample. The matching weights are included to estimate the coefficients. Column 2 and Column 4 in Table 5 summarize the estimations with the matched sample. The reported standard errors in the parentheses are clustered- robust standard errors. Column 1 and Column 3 are OLS estimates with unmatched sample and they are included as comparisons. When the dependent variable is the depression level, the estimated coefficient of treatment variable now becomes significant at 5% level with the matched sample and there's a slight increase in the magnitude. This indicates that living in rural area lead to severer depression problems. When the dependent variable is whether the individual takes prescription for depression, the significance and the magnitude have not changed. The estimation obtained from matched sample still predicts a 3% increase in probability of taking prescription when individuals move to rural area.

Table 5. Estimated Effects of Living in Rural Areas With Matched Sample

	Dependent variable:			
	depression		prescription	
	(1)	(2)	(3)	(4)
Rural	0.028*	0.039**	0.030***	0.030***
	(0.015)	(0.018)	(0.007)	(0.008)
Constant	2.315***	2.239***	0.195***	0.118*
	(0.120)	(0.166)	(0.052)	(0.070)
Controls	Yes	Yes	Yes	Yes
Matched Sample	No	Yes	No	Yes
Observations	9,398	9,398	9,398	9,398
Adjusted R^2	0.190	0.182	0.082	0.089
Note:	*p<0.1; **p<0.05; ***p<0.01			

4. Conclusion

This paper uses the NHIS 2019 survey data to identify the effect of living in rural area on the depression level. Although previous literatures find that living in urban area is associated with higher risk of depression and higher usage of prescription, I find that living in rural area is associated with severer depression and higher usage of prescription by employing a propensity score matching method.

However, several concerns exist with this analysis. First, comparing with previous liter- atures, this paper uses cross sectional data from a specific year instead of panel data over several years. The number of observations is not that large as those literatures studying the same question. Panel data also contains more information than cross sectional data, such as the time effect. The estimates might be more accurate if panel data is used. Second, the two dependent variables of interest are categorical variables, and one of them has ordinal information. But this paper only use linear regression model to study the effect of place of residence on the dependent variables. For the ordinal dependent variable, it's hard to interpret the results. For the binary dependent variable, the linear probability model might predict probability outside the range between zero and one. A logit/probit framework should be considered in the future work. Another concern is the potential failure of conditional in- dependence assumption. Although the propensity score matching method can help reduce the bias, it still requires the conditional independence assumption to hold in the applica- tion. But this assumption is often untestable and it is likely that there still exist omitted factors that simultaneously affect the treatment variable and the outcome variable in the application. Therefore, the structural model for the risk of depression should be studied in the future work, or use a different framework which does not rely heavily on conditional independence assumption.

References

[1] Ayuso-Mateos, J. L., Vázquez-Barquero, J. L., Dowrick, C., Lehtinen, V., Dalgard, O. S., Casey, P., Wilkinson, C., Lasa, L., Page, H., Dunn, G., Wilkinson, G., & ODIN Group (2001). Depressive disorders

in Europe: prevalence figures from the ODIN study. The British journal of psychiatry : the journal of mental science, 179, 308–316. https://doi.org/10.1192/bjp.179.4.308

[2] Patten SB, Wang JL, Williams JV, et al. Descriptive Epidemiology of Major Depression in Canada. The Canadian Journal of Psychiatry. 2006;51(2):84-90. doi:10.1177/070674370605100204

[3] CDC.. About Mental Health. Retrieved from https://www.cdc.gov/mentalhealth/learn/index. (2021, June 28)

[4] Layard, R.. The Economics of Mental Health. (2017, January 1). Retrieved from https://wol.iza.org/uploads/articles/321/pdfs/economics-of-mental-health.pdf

[5] Knapp M, Wong G. Economics and mental health: the current scenario. World Psychiatry. 2020 Feb;19(1):3-14. doi: 10.1002/wps.20692. PMID: 31922693; PMCID: PMC6953559

[6] J. Tom Mueller, Kathryn McConnell, Berne Burow, et al. Impacts of the COVID-19 pandemic on rural America. PNAS. 2020

[7] Van Os, J. (2004). Does the urban environment cause psychosis? British Journal of Psychiatry, 184(4), 287-288. doi:10.1192/bjp.184.4.287

[8] Maslow, A., & Lewis, K. J. (1987). Maslow's hierarchy of needs. Salenger Incorporated, 14(17), 987-990.

[9] Crowell, Bradford A., Linda K. George, Dan Blazer, and Richard Landerman. 1986. “Psy- chosocial Risk Factors and Urban/Rural Differences in the Prevalence of Major Depres- sion.” British Journal of Psychiatry 149 (3):307–314

[10] McKenzie, Karen, Aja Murray, and Tom Booth. 2013. “Do urban environments increase the risk of anxiety, depression and psychosis? An epidemiological study.” Journal of Affective Disorders 150 (3):1019–1024

[11] Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal In- ference.” Political Analysis 15 (3):199–236

.