Women's Resources and Their Risks of Intimate Partner Violence: Evidence from the Democratic Republic of Congo

Qiwei Lin, UMID: 76432810

April 26 2020

Contents

1 Introduction **3**

2 Data **4**

3 Method **7**

3.1 Methods to Examine Association Between Two Discrete Variables 7

3.2 Methods to Examine Association Between one Continuous and one Discrete Variable 9

3.3 Methods to Establish Confidence Interval for Coefficients in Logistic Regression . 9

4 Simulation Study **10**

4.1 Simulation Study of χ^2 Test, G^2 Test, and Difference-in-Proportion 10

4.2 Simulation Study of χ^2 Test, G^2 Test and M^2 Test 11

4.3 Simulation Study for Logistic Regression and Bootstrap Confidence Interval 11

5 Analysis **13**

5.1 Permutation Test on Women's Employment Status and Experience of IPV 13

5.2 Permutation Test on Women's Property Ownership and Experience of IPV 13

5.3 Permutation Test on Women's Household Wealth and Experience of IPV 13

5.4 Logistic Regression 14

6 Discussion **15**

Appendix **17**

Appendix I 17

Appendix II 18

Works Cited **20**

List of Tables

1 Sample Characteristics, 12448 Women in DRC in 2013 6

2 Simulated Operating Characteristics of the Tests Based on 10,000 Monte Carlo Samples 10

3 Simulated Size and Power of the Tests Based on 10,000 Monte Carlo Samples . . . 11

4 Logistic Regression Coefficients 14

List of Figures

1 99% Confidence Interval for The Bias of Coefficient Estimates 12

2 Employment Status vs Violence 18

3 Property Ownership vs Violence 18

4 Household Wealth Index vs Violence 19

5 Education vs Violence 19

1 Introduction

Violence against women is a violation of human rights and among all forms of violence, intimate partner violence (IPV) is considered to be the most common form (Devries et al.). This issue is particularly salient in developing countries such as the Democratic Republic of Congo (DRC), which is estimated to have the highest rate of reported violence (Cools and Kotsadam). IPV adversely affects women's health and also creates negative externalities such as fear of abuse and psychological stress for those who witness violence. The severity and negative consequences of IPV highlight the need to understand factors associated with the prevalence of IPV and the importance of proposing policy solutions to this issue. In this paper, I used permutation tests and logistic regression to explore whether in DRC women's probability of experiencing IPV is associated with the following four individual-level characteristics of women: employment status, property ownership, household wealth, years of education.

Literature in international development has proposed various hypotheses that explain why four characteristics above might be associated with the prevalence of IPV. The resource theory in international development states that resources possessed by individual women are empowering and protective against IPV (Jewkes). These resources include both physical resources such as income, wealth, and human resources like education, social network, abilities to use information and resources in society.

One channel through which women's resources reduce their probability of experiencing IPV is reducing the economic stress of the household. Literature shows that economic stress tends to lead to domestic violence (Martin et al.). Women who are employed and paid might therefore face a lower risk because their income can help reduce the household economic stress. Similarly, Women from relatively wealthier households might also face a lower risk of IPV as a result of lower economic stress. In addition to stress-reduction mechanism hypothesized in the resource theory, employed women may also reduce their chance of being abused at home through an exposure-reduction mechanism (Chin). As a result of being employed, women spend less time with perpetrators at home and therefore reduce their risks of experiencing IPV.

Another channel through which resources become protective against IPV is improving their outside options such as divorce and increasing women's bargaining power within the household. Owning properties such as a house or lands might be protective against IPV as this property ownership reduces women's economic dependence on partners and protects women from being homeless if they choose to divorce. As a result, women's bargaining power at home increases as they can use the threat to divorce as a deterrence to perpetrators at home to reduce their risks of experiencing IPV.

Women's education can also be protective against IPV through the similar mechanism. Women's education facilitates empowerment by expanding their social networks, boosting self-confidence, and improving abilities to use information and resources available in society, and may also translate into wealth (Jewkes). As a result of empowerment, women become less dependent and more capable of using the divorce threat to reduce their probability of experiencing IPV.

The structure of the paper is as follows. The Data section introduces the dataset, the definition of variables of interests and relevant data processing methods I applied on the data. Method section briefly provides background on the hypothesis testing methods and statistical models I used in the analysis. Simulation section presents the assessment of those methods based on Monte Carlo simulations. Analysis section summarizes the main result and presents the substantive interpretation. Discussion section will review the discussion, explore the implication of the result and point out some limitations that future study can address.

2 Data

The data for this study comes from the Demographic and Health Surveys (DHS). It is a series of nationally representative surveys taken by United States Agency of International Development in developing countries. For DRC, the most recent survey was conducted in 2013 and the data is representative at both national and province level (see (DHS, "DHS Methodology") for details on data and sampling methodology). This 2013 survey dataset contains individual-level information for women in DRC who are currently in union and aged 15 to 49.

Three types of intimate partner violence are measured in the DHS dataset, emotional violence, sexual violence, and physical violence. Each type of violence is recorded as a binary indicator that uses 1 to denote past experience of violence and 0 otherwise. Survey takers asked a series of behavioral questions to assess whether a respondent experienced a specific type of domestic violence (in the past 12 month) and the value on an indicator variable will be 1 if a respondent answer yes on at least one behavioral question (Appendix I presents the full list of behavioral questions used in the assessment). I combined three indicators to construct the variable, *violence*, that uses 1 to denote that a respondent experienced at least one type of domestic violence 12 months prior to the survey.

The DHS dataset also includes information on several individual-level characteristics that measure women's resource endowment and that might capture differences between victims and non-victims. The following eight variables are used in my analysis: *employ_paid*, *owning_property*, *wealth_index*, *education*, *partner_edu*, *age* and *urban*. *Employ_paid* is a binary indicator that uses 1 to denote that the respondent is employed and receives either cash or in-kind payments. *owning_property* is also a binary indicator. It uses 1 to denote that a respondent owns either house or lands. *Wealth_index* is an ordinal variable that measures a respondent's household wealth. It has five levels: poorest, poorer, middle, richer, and richest. *Education* is a discrete, numerical variable measuring total years of education a respondent receives. The following three variables (*partner_edu*, *urban*, *age*) will be used as control variables in the logistic regression model. *Partner_edu* is a discrete, numerical variable measuring total years of education a respondent's partner receives. *Age* is a discrete, numerical variable that measures the respondent's age. *Urban* is a binary variable measuring a woman's residency. If a woman lives in an urban area, she will have *urban* = 1. Table 1 presents the summary statistics for these variables. Exploratory plots of $violence$ and four variables of interest can be found in Appendix II.

DHS data, however, has a missing data issue. Table 1 shows that *employ_paid*, *owning_property*

Table 1: Sample Characteristics, 12448 Women in DRC in 2013

	% or Mean, M	SD	proportion of missing (%)
violence, %	55.7	0.50	58.91
emo_vio, %	35.4	0.48	58.89
sex_vio, %	24.9	0.43	58.88
phy_vio, %	45.2	0.50	58.90
employ_paid, %	78.3	0.41	0.06
owning_property, %	62.4	0.48	0.03
education, M	5.06	3.96	0
partner_edu, M	8.47	4.19	2.43
age, M	30.78	8.56	0
urban, %	30.4	0.46	0
wealth_index			
poorest, %	24.80	0.43	0
poorer, %	22.69	0.42	0
middle, %	21.02	0.41	0
richer, %	17.46	0.38	0
richest, %	14.03	0.35	0

Note: For variables with missing values, summary statistics are calculated based on available cases.

and *partner_edu* have a relatively small proportion of missingness and I treated these missing values as Missing Completely At Random (MCAR). The key outcome variable *violence*, however, has a large proportion of missing values (about 60%) due to missingness on three indicator variables. This pattern of missingness is likely to be Missing Not At Random (MNAR). It is possible that women who experienced domestic violence are more likely to fail to report their experience of violence due to their fear for perpetrators at home. If I delete all observations with at least one missing value, the complete dataset I obtain is not a representative sample of the population.

To account for this missing data problem, I prepared two datasets for two different scenarios and compared results obtained from the analysis of each dataset. In the first scenario, I assumed all missingness in the *violence* variable are MCAR and drop all missing values. Under the MCAR assumption, the complete dataset is still representative of the population even if I delete all missing values. The second scenario is the extreme scenario in which I assumed all women who failed to report their experience of domestic violence are actual victims and replaced all missing values with

1. This dataset is arguably more representative than the complete dataset in the first scenario, but it might not be representative to the true population.

3 Method

3.1 Methods to Examine Association Between Two Discrete Variables

Women who are employed and paid and those who did not constitute two potentially different samples. Similarly, women who own properties or not also form two samples. I performed permutation tests on employment status and chance of experiencing IPV, and on property ownership and IPV, and drew inference on whether two groups are significantly different.

Permutation test is a non-parametric test method relying on resampling without replacement (Rizzo). Under the null hypothesis that two groups of women are not different in experiencing IPV, I permuted women's employment status and property ownership and computed the test statistics T 1000 times to obtain the null distribution of the test statistics T. I then sustain or reject the null hypothesis by checking how extreme the observed test statistics is under the null distribution. The advantage of permutation tests is that it makes no assumption about the null distribution of the test statistics T and uses data to generate that distribution instead.

There are three candidate tests for the permutation tests on the association between two binary variables. Difference-in-proportion test, Chi-squared test and Likelihood-ratio chi-squared test (also known as G^2 test). Difference-in-proportion test uses the test statistics $\hat{p}_x - \hat{p}_y$, where $\hat{p}_x$ and $\hat{p}_y$ are the observed sample proportions. When the number of observations in both variables (n_x and n_y) are relatively large, we can use the Central Limit Theorem to approximate the distribution of $\hat{p}_x$, $\hat{p}_y$ and $\hat{p}_x - \hat{p}_y$ (Navidi). We then have the following results:

$$\hat{p}_x \sim N(p_x, \frac{p_x(1-p_x)}{n_x}) \quad , \quad \hat{p}_y \sim N(p_y, \frac{p_y(1-p_y)}{n_y}) \tag{1}$$

$$\hat{p}_x - \hat{p}_y \quad \sim \quad N(p_x - p_y, \frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}) \tag{2}$$

For two discrete variables X and Y with I and J categories separately, the test statistics used in Chi-squared test is:

$$\chi^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \tag{3}$$

and the test statistics in G^2 test is:

$$G^2 = 2 \sum n_{ij} \log(\frac{n_{ij}}{\mu_{ij}}) \tag{4}$$

Under a pre-specified null hypothesis, cell probabilities in the contingency table of X and Y equal some fixed values π_{ij}. π_{ij} are usually 0.5 when both X and Y are binary variables. $\mu_{ij} = n\pi_{ij}$ are expected frequencies under the null hypothesis and n is the number of observation in the whole dataset. n_{ij} is the observed cell counts in the table. Both χ^2 and G^2 follow a chi-squared distribution with degree of freedom = $(I-1)(J-1)$ (Agresti).

The variable *wealth_index* is an ordinal variable with five levels and associated integer scores from 1 to 5. I can still use a permutation test to examine the association between a woman's household wealth and her chance of experiencing IPV. Difference-in-proportion test is not applicable, but both Chi-squared test and G^2-test can still be used to assess whether there is a dependence between these two variables. However, the disadvantage of using these two tests is that both tests ignore the ordering information. Test statistics that use ordering information by treating ordinal variables as numerical variables are more appropriate and superior in the power (Agresti).

A candidate test in this category is M^2 test with the following test statistics:

$$\begin{aligned} M^2 &= (n-1)r^2 \qquad (5) \\ r &= \frac{\sum_{i,j}(u_i - \bar{u})(v_j - \bar{v})p_{ij}}{\sqrt{[\sum_i (u_i - \bar{u})^2 p_{i+}]}\sqrt{[\sum_j (v_j - \bar{v})^2 p_{j+}]}} \qquad (6) \end{aligned}$$

Assume that we organize two discrete variables of interests X and Y into an $I \times J$ contingency table. Let $u_1 \leq u_2 \leq \cdots \leq u_I$ be the numerical scores for the row variable, and let $v_1 \leq v_2 \leq \cdots \leq v_J$

be the scores for the column variable. $\bar{u} = \sum_i u_i p_{i+}$ is the sample mean of the row score while $\bar{v} = \sum_j u_j p_{+j}$ denotes the sample mean of the column scores. p_{ij} is the relative frequency of observations in this group. r is basically the correlation between X and Y. When the number of observations grows large, M^2 approximately follows a chi-squared distribution with degree of freedom 1 (Agresti).

3.2 Methods to Examine Association Between one Continuous and one Discrete Variable

To examine the association between a woman's probability of being abused and her years of education, I fitted the following logistic regression model:

$$\begin{aligned} \text{logit(Pr(violence = 1))} = \beta_0 &+ \beta_1 employ\_paid + \beta_2 owning\_property + \beta_3 wealth\_index \\ &+ \beta_4 education + \beta_5 partner\_edu + \beta_6 urban + \beta_7 age \end{aligned} \tag{7}$$

This logistic regression model models the probability of experiencing IPV as a function of the linear combination of a woman's individual-level characteristics using the logistic function $\ln(\frac{p}{1-p})$ as its canonical link function with p = the probability of experiencing IPV for a woman. β_4 will shed light on the association between a woman's probability of experiencing IPV and the years of education she received when we control for other individual characteristics.

3.3 Methods to Establish Confidence Interval for Coefficients in Logistic Regression

I used the non-parametric bootstrap method to calculate the 95% confidence interval for the coefficient in the logistic regression model above. I sampled observations from the dataset with replacement to obtain one bootstrap dataset and fitted the logistic regression on this dataset. This process is repeated 1000 times and I then obtained 1000 bootstrap estimates for each coefficient.

These bootstrap estimates for coefficients constitute the sampling distribution of coefficients. I then took the 2.5% and 97.5% quantile of the sampling distributions and uses those numbers to construct percentile intervals. Those percentile intervals are the 95% bootstrap confidence interval for the coefficients (see Rizzo for detailed discussions about bootstrap methods and percentile interval).

4 Simulation Study

4.1 Simulation Study of χ^2 Test, G^2 Test, and Difference-in-Proportion

In this section, I used 10000 Monte Carlo simulations to assess the Type I error rate (also known as the size) and the power (the probability of rejecting a null hypothesis when it is false) of χ^2 test, G^2 test, and Difference-in-Proportion. To assess the size of a test, I generated two sets of Bernoulli random variables with no association or difference in each Monte Carlo sample, applied each test method, and calculated the proportion of results that sustain the null hypothesis of no association/difference. To assess the power of a test, I generated two sets of Bernoulli random variables with dependence and difference, applied the tests again and calculated the proportion of results that reject the null hypothesis of no association/difference.

Table 2 summarizes the estimates and the 95% confidence interval of the size and power for these three tests. All three tests have a reasonably well size with the commonly used $\alpha = 0.05$ included in its 95% confidence interval. In terms of the power, G^2 test seems to slightly outperform other two tests with its power higher than other two tests by at least 0.01. I thus decided to use G^2 test as the test method in the permutation test to examine the association between women's employment status and experience of IPV, and between their property ownership and experience of IPV.

Table 2: Simulated Operating Characteristics of the Tests Based on 10,000 Monte Carlo Samples

	Size		Power	
	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
χ^2 test	0.044	(0.040, 0.048)	0.882	(0.876, 0.888)
Diff-in-Prop	0.049	(0.045, 0.053)	0.872	(0.866, 0.879)
G^2 test	0.049	(0.044, 0.053)	0.892	(0.886, 0.898)

4.2 Simulation Study of χ^2 Test, G^2 Test and M^2 Test

In this section, I used 10000 Monte Carlo simulations to assess the size and the power of χ^2 test, G^2 test and M^2 test. To assess the size of a test, I generated one set of Bernoulli random variables and one set of five-level ordinal variables (with assigned integer scores between 1 to 5) with no association or difference in each Monte Carlo sample, applied each test method, and calculated the proportion of results that sustain the null hypothesis of no association. To assess the power of a test, I generated the set of Bernoulli random variables conditional on the ordinal variables and calculated the proportion of results that reject the null hypothesis of no association.

Table 3: Simulated Size and Power of the Tests Based on 10,000 Monte Carlo Samples

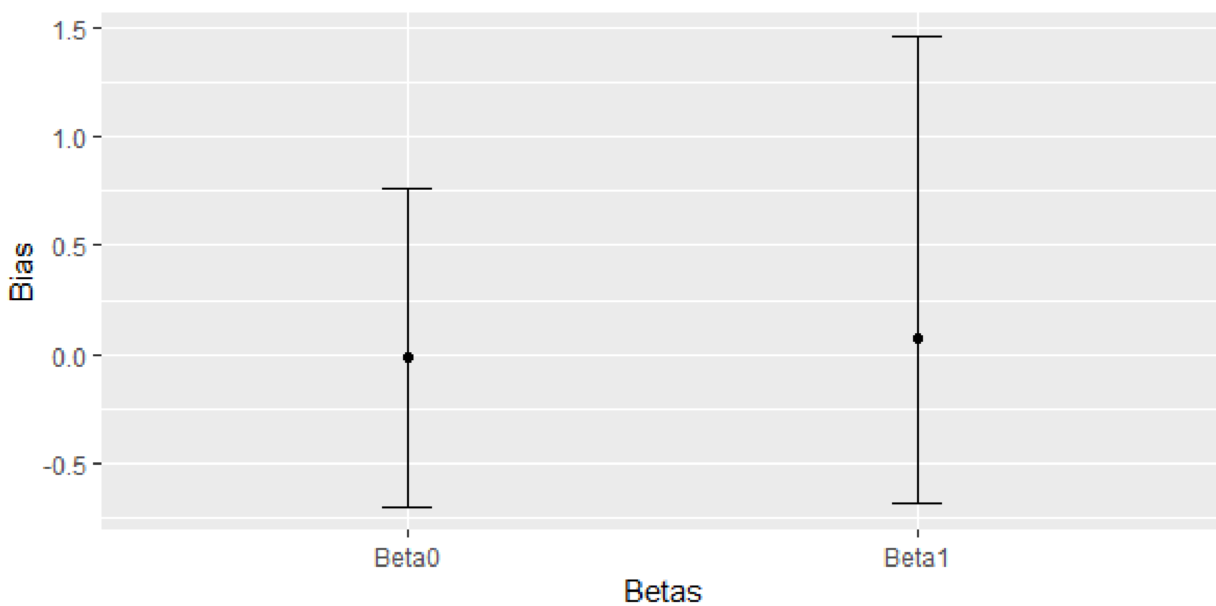
	Size		Power	
	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
χ^2 test	0.047	(0.043, 0.052)	0.844	(0.837, 0.851)
G^2 test	0.050	(0.046, 0.055)	0.833	(0.826, 0.840)
M^2 test	0.050	(0.046, 0.054)	0.952	(0.947, 0.956)

Table 3 summarizes the estimates and the 95% confidence interval (CI) of the size and power for these two test statistics. Three tests have satisfactory Type I error rate, all reaching the nominal level. M^2, however, outperforms χ^2 test and G^2 test in terms of the power. The power of M^2 is estimated to be 0.952 and is higher than that of χ^2 by about 0.11, and than that of G-test by about 0.12. This might be explained by that M^2 test uses the ordinal information in the data while χ^2 test and G-test do not. It seems that M^2 is a better test statistics for the permutation test on women's household wealth and their experience of IPV.

4.3 Simulation Study for Logistic Regression and Bootstrap Confidence Interval

In the third simulation study, I evaluated the bias of coefficient estimates from a logistic regression model and the coverage rate of bootstrap percentile confidence interval for the coefficients. I simulated 2000 Monte Carlo samples in which Pr($Y = 1$) is conditional on $\frac{1}{1+exp\{-\beta_0-\beta_1 X\}}$. I then estimated a logistic regression model on each sample. Then, I calculated the bias of each coefficient estimate in each sample. Figure 1 shows the 99% confidence interval for the bias of two coefficients. Both intervals include 0, which indicates that we cannot reject the null hypothesis that the coefficient estimates from logistic regression are unbiased at $\alpha = 0.01$. This result provides a justification for using logistic regression to estimate the association between the probability of IPV and female characteristics in DRC.

Figure 1: 99% Confidence Interval for The Bias of Coefficient Estimates



To assess the coverage rate of bootstrap confidence interval, I generated 100 95% bootstrap confidence intervals and calculated the coverage rate of intervals for the true parameters. The result shows that for both intercept and the coefficient, the coverage rate is 100%. Limited by the computing power and run time, I only generated 100 confidence intervals, which might contribute to this performance that is much better than the 95% nominal level I expect.

5 Analysis

5.1 Permutation Test on Women's Employment Status and Experience of IPV

In the complete case analysis, the observed test statistics is G^2 = 10.652 with an associated p-value 0.002. In the extreme case analysis, the observed test statistics is G^2 = 0.727 with an associated p-value 0.824. Women's employment status seems to have an association with their probability of experiencing violence in the complete-case scenario where the sample is arguably less representative due to the potential missing not at random problem. The result of permutation test only partially supports that women's employment status is associated with their experience of IPV.

5.2 Permutation Test on Women's Property Ownership and Experience of IPV

In the complete case analysis, the observed test statistics is G^2 = 12.579 with an associated p-value 0.002. In the extreme case analysis, he observed test statistics is G = 29.679^2 with an associated p-value 0. The result of permutation test indicates that we can reject the null hypothesis that women's property ownership has no association with their experience of IPV under both complete-case and extreme-case scenarios.

5.3 Permutation Test on Women's Household Wealth and Experience of IPV

In the complete case analysis, the observed test statistics is M^2 = 3.384 with an associated p-value 0.138. In the extreme case analysis, The observed test statistics is M^2 = 2.204 with an associated p-value 0.29. The results of permutation test does not lend support to the hypothesis that women's household wealth is associated with their experience of IPV.

5.4 Logistic Regression

The main results of logistic regression are summarized in Table 4. We can see that in the

Table 4: Logistic Regression Coefficients

	Complete Case (n = 5010)		Extreme Case (n = 12141)	
	Coefficient	95% Bootstrap CI	Coefficient	95% Bootstrap CI
employ_paid	0.282	(0.139, 0.421)	0.103	(-0.0002, 0.219)
education	-0.008	(-0.029, 0.010)	-0.009	(-0.024, 0.007)
owning_property	-0.241	(-0.366, -0.116)	-0.272	(-0.384, -0.163)
wealth_index	-0.069	(-0.130, -0.013)	-0.031	(-0.078, 0.014)
partner_edu	-0.009	(-0.027, 0.008)	0.005	(-0.009, 0.019)
urban	0.241	(0.076, 0.410)	0.227	(0.089, 0.368)
age	-0.003	(-0.011, 0.004)	0.008	(0.002, 0.013)
constant	0.491	(0.229, 0.765)	1.374	(1.143, 1.602)

[a] The 95% confidence intervals are based on 1000 bootstrap estimates and Percentile Interval

complete case analysis, women who are employed and paid are more likely to experience IPV when controlling for other individual characteristics. This association is significant at $\alpha = 0.05$. If we hold other variables at the median level, the probability of experiencing IPV for employed women who receive payment is 7% higher than that of other women. This contradicts the resource theory and and the hypothesis that women who are employed and paid possess more resources and face a lower risk of IPV. In the extreme case analysis, women who are employed and receive payments still face a higher risk but the association is not significant at $\alpha = 0.05$ level. Women's household wealth is negative correlated with their chance of experiencing IPV. The probability of experiencing IPV decreases by 1.72% when women's households move from the poorest level to the poorer level when we hold other characteristics at the median level. However, this relation is only statistically significant in the complete case analysis where the sample is less representative due to missingness. In the extreme case analysis where the sample is arguably more representative, we cannot reject the null hypothesis that women's household wealth has no correlation with their probability of being abused at $\alpha = 0.05$ level. The result only partially supports the hypothesis that women from a wealthier household are less likely to be abused.

Women's education is negatively correlated with the probability of experiencing IPV in both complete- and extreme-case. These associations, however, are not significant at $\alpha = 0.05$ level in both cases. The data does not lend a strong support for the reasoning that education is a protective resource against IPV.

We can reject the null hypothesis that women's property ownership is not correlated with their probability of experiencing IPV at $\alpha = 0.05$ level. From Table 4, property ownership is negatively correlated with women's probability of experiencing IPV in both cases. I calculated the reduction on the probability of experiencing violence associated with owning property for a 30-year-old woman from a rural house with middle wealth who is employed and paid, receives median years of education, and has a partner who also receives an median level of education. In the complete case analysis, owning a house or lands is associated with a reduction of 5.95% on the probability of experiencing IPV. In the extreme case analysis, the associated reduction is 4.11%. It seems that owning property has protective impacts on women in DRC.

6 Discussion

In this paper, I examined the association between variables that measure women's physical and human resources and their probability of experiencing IPV with data from the Democratic Republic of Congo in 2013. Due to a large proportion of missingness in the data, I prepared two datasets for analysis: the complete-case dataset with all missing values excluded and the extreme-case dataset with all women who failed to report their experience of IPV treated as victims of IPV. Results of simulation studies recommend G^2 test and M^2 test as test statistics in permutation tests since these two tests have a higher power. Simulation studies also provide justifications for using logistic regression to estimate the coefficients of interests for its unbiasness and using the non-parametric bootstrap percentile interval to establish 95% confidence interval for coefficient estimates for its satisfactory coverage rate.

The main finding for this paper is that among four resources that are hypothesized to be protective, only women's property ownership seems to be protective against IPV in both complete-case and extreme-case analysis. The probability of experiencing IPV decreases by 5.95% and 4.11% respectively for a woman who owns either houses or lands when we hold characteristics at the median level. Women's household wealth only shows protective effects in the complete-case analysis. The policy implication of these results is that governments might reduce IPV through women's empowering programs that promote women's property ownership or cash-transfer programs that reduce households' economic stress.

Although this paper found some useful insights, some limitations are also worth noting. This paper identifies some correlations between women's characteristics that measure their resources endowment and their probability of being abused, but these correlations do not imply causal relationships between women's resources and their chance of being abused. We cannot rule out the reverse causality that women who have less abusive partners are more likely to own properties. Future study can use government interventions such as cash-transfer programs, development programs that promote women's employment and education as randomized experiments to examine whether there are causal effects of changes in women's resource endowment on their probability of experiencing IPV.

Another issue is that analysis in this paper does not account for the group structure for women in DRC. The effect of women's resources on their probability of being abused might differ across provinces of DRC since the macro-level context of a province might affect women differently. It is possible that the protective effect of owning property against IPV is only robust in provinces where the discriminatory social norm that accepts violence against women is weak. Future analysis can examine whether the relations identified in this paper are still robust when the macro-level, contextual effects of provinces are accounted for in models like multilevel models. The third issue is about the missing data problem. In this study, I used the complete-case dataset obtained by using the list-wise deletion method and the extreme-case dataset obtained by treating all missing on experience of IPV as ”Yes”. These two datasets might not be representative to the true population. Future study can instead use multiple imputations to obtain a sample that might be more representative to the true population and re-examine the relationship found in this paper.

Appendix

The list below is based on the Demographic and Health Surveys Standard Recode book (see (DHS, "Standard Recode Manual for DHS7") for detailed information).

Appendix I: Behavioral Questions

Behavioral Questions for Emotional Violence

- Did your spouse ever humiliate you?
- Did your spouse ever threaten you with harm?
- Did your spouse ever insult you or make you feel bad?

Behavioral Questions for Physical Violence

- Did your spouse ever push, shake or throw something?
- Did your spouse ever ever slap you?
- Did your spouse ever punch you with fist or something harmful?
- Did your spouse ever kick or drag you?
- Did your spouse ever try to strangle or burn you?
- Did your spouse ever threaten you with knife/gun or other weapon?
- Did your spouse ever attack you with knife/gun or other weapon?
- Did your spouse ever twist you arm or pull your hair?

Behavioral Questions for Sexual Violence

- Did your spouse ever physically force sex when not wanted?
- Did your spouse ever force other sexual acts when not wanted?

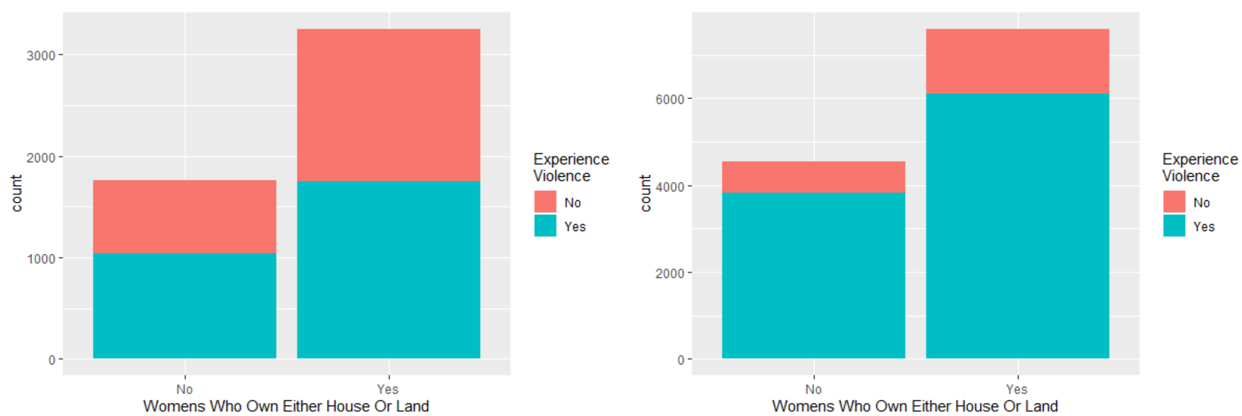
Appendix II: Exploratory Analysis

Figure 2: Employment Status vs Violence

(a) Complete-Case Dataset

(b) Extreme-Case Dataset

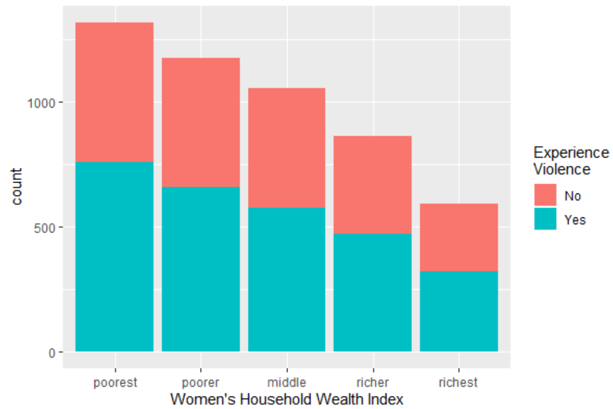
Figure 3: Property Ownership vs Violence



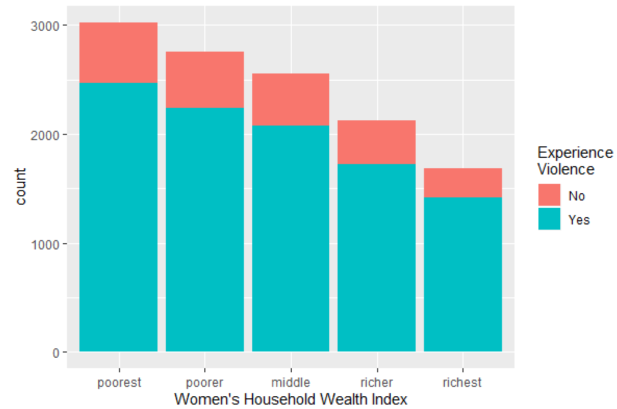
(a) Complete-Case Dataset

(b) Extreme-Case Dataset

Figure 4: Household Wealth Index vs Violence

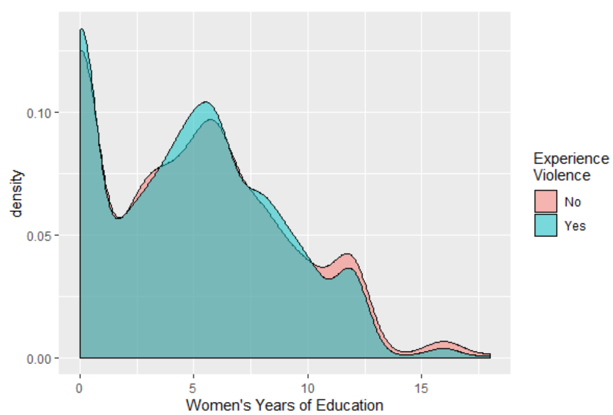


(a) Complete-Case Dataset



(b) Extreme-Case Dataset

Figure 5: Education vs Violence



(a) Complete-Case Dataset

(b) Extreme-Case Dataset

Works Cited

Agresti, Alan. *An introduction to categorical data analysis*. 2nd ed. Wiley-Interscience, 2007. Print.

Chin, Yoo-Mi. "Backlash, Bargaining, or Exposure Reduction?: Women's Working Status and Physical Spousal Violence in India." *Journal of Population Economics* 25.1 (Aug. 2012): 175–200. Web.

Cools, Sara and Andreas Kotsadam. "Resources and Intimate Partner Violence in Sub-Saharan Africa". *World Development* 95 (2017): 211–230. Web.

Devries, K. M., et al. "The Global Prevalence of Intimate Partner Violence Against Women". *Science* 340.6140 (2013): 1527–1528. eprint: https://science.sciencemag.org/content/340/6140/1527.full.pdf. Web.

DHS. "DHS Methodology". 2020. Accessed: 2020-04-25. https://dhsprogram.com/What-We-Do/Survey-Types/DHS-Methodology.cfm#CP_JUMP_16156.

—–."Standard Recode Manual for DHS7". 2018. Accessed: 2020-04-25. https://dhsprogram.com/pubs/pdf/DHSG4/Recode7_DHS_10Sep2018_DHSG4.pdf.

Jewkes, Rachel. "Intimate partner violence: causes and prevention". *The Lancet* 359.9315 (2002): 1423–1429. Web.

Martin, Sandra L., et al. "Domestic Violence in Northern India". *American Journal of Epidemiology* 150.4 (Aug. 1999): 417–426. eprint: https://academic.oup.com/aje/article-pdf/150/4/417/227849/150-4-417.pdf. Web.

Navidi, William Cyrus. *Statistics for engineers and scientists*. Fourth edition. McGraw-Hill Education, 2015. Print.

Rizzo, Maria L. *Statistical computing with R*. Second Edition. CRC Press, Taylor Francis Group, 2019. Print. Chapman vand Hall/ CRC the R series.