ReThinking Emergency Rental Allocation

Appendix C. Econometric models descriptors

ReThinkers Regina Isabel Medina Rosales, Sanhitha Cherukupally, Tasaduq Soomro, Carlos Enrique Huapaya

October 16th, 2022

Contents

Being at risk of eviction is a result of several factors. Recognizing the multicasual nature of this social phenomenon, the *ReThinkers* team designed a quantitative strategy sensitive to both personal-level factors and county-level factors.

Datasets

- Household Pulse Survey (HSP)
- Census data
- State eviction laws data sets
- Treasury Emergency Rental Assistance (ERA) Dashboard

Software

For estimation of econometric modeling we used the following R-Packages:

- A grammar of data manipulation (`dplyr` package[@wickham2015dplyr])
- Elegant graphics for data analysis (`ggplot2` package[@wickham2011ggplot2])
- Analysis and Presentation of Social Scientific Data (`jtools` package[@long2017package])
- Econometric modeling (`stats` package[@R-core])
- GAM and Generalized Ridge Regression for R (`mgcv` package[@wood2012mgcv])

`dplyr`

A grammar of data manipulation, `dplyr` package, was built by Hadley Wickham. It "provides a consistent set of verbs that help you solve the most common data manipulation challenges".[@wickham2015dplyr] We used `dplyr` for the initial data processing in order to have clean data suited for the modeling and visualization processes.

`ggplot2`

All the visualizations in this report were built using the `ggplot2` package, created by Hadley Wickham, which is "a system for declaratively creating graphics, based on The Grammar of Graphics" by Leland Wilkinson[@wilkinson2012grammar].

`epitools`

We also took advantage of the `epitools` package, a set of "basic tools for applied epidemiolgy".[@aragon2012epitools] We used the package to estimate the observed variation of homicide data.

`lme4`

All the models from the fist section were buil using the `lme4` package, created by Ben Bolker. It is designed to fit Linear, Generalized Linear, And Nonlinear Mixed models, providing different functions such as `lmer()`, `glmer()`, and `nlmer()`.[@bates2007lme4] Since we are modeling counts and rates (assuming a Poisson distribuiton for the dependent variable), we used the `glmer()` function, fitted for Generalized Linear Mixed Models (GLMM).

`jtools`

Jacob A. Long wrote the `jtools` package as a set of tools "for the purpose of more efficiently understanding and sharing the results of regression analyses.[@long2017package] For our analysis, we used the `summ()` function to present the results of the `glmer()` models. The `summ()` command"prints output for a regression model in a fashin similar to `summary`, but formatted differently with more options", which include including data description, goodness of fit measurements, as well as fixed and random effects coefficients.[@long2017package]

`merTools`

The `merTools` package, Tools for Analyzing Mixed Effect Regression Models, "provides methods for extracting results from mixed-effect model objects fit with the `lme4` package" and "allows construction of prediction intervals efficiently from large scales linear and generalized linear mixed-effects models".[@knowles2016mertools] From this package, we used the `predictInterval` command, wich provides the confidence intervals for the random effects coefficients using bootstrapping techniques.

`mgcv`

Finally, we use the `mgcv` package, Mixed GAM Computation Vehicle with Automatic Smoothness Estimation, by Simon Wood to perform final modeling with the smoothnes transformations.[@wood2012mgcv] In particular, we used the `gamm()` function for modeling Generalized Additive Mixed Models with the smoothness (achived through the use of splines) applied to the *time* variable.

Probit model for personal-level factors

The following snipet code shows the probit model specification.

```r
# Model
m1 <- glm(
  eviction_likelihood ~ age + female + white + black + asian +
    hispanic + non_het + trans + single + kids + college +
    unemployment_month + military + disability_phys + disability_mental +
    rent_assistance + rent_increase + rent_delay_months + (single*kids),
    family = binomial(link = "probit"),
    weights = df_clean$pweight,
    data = df_clean,
    # na.rm = T
    )
```

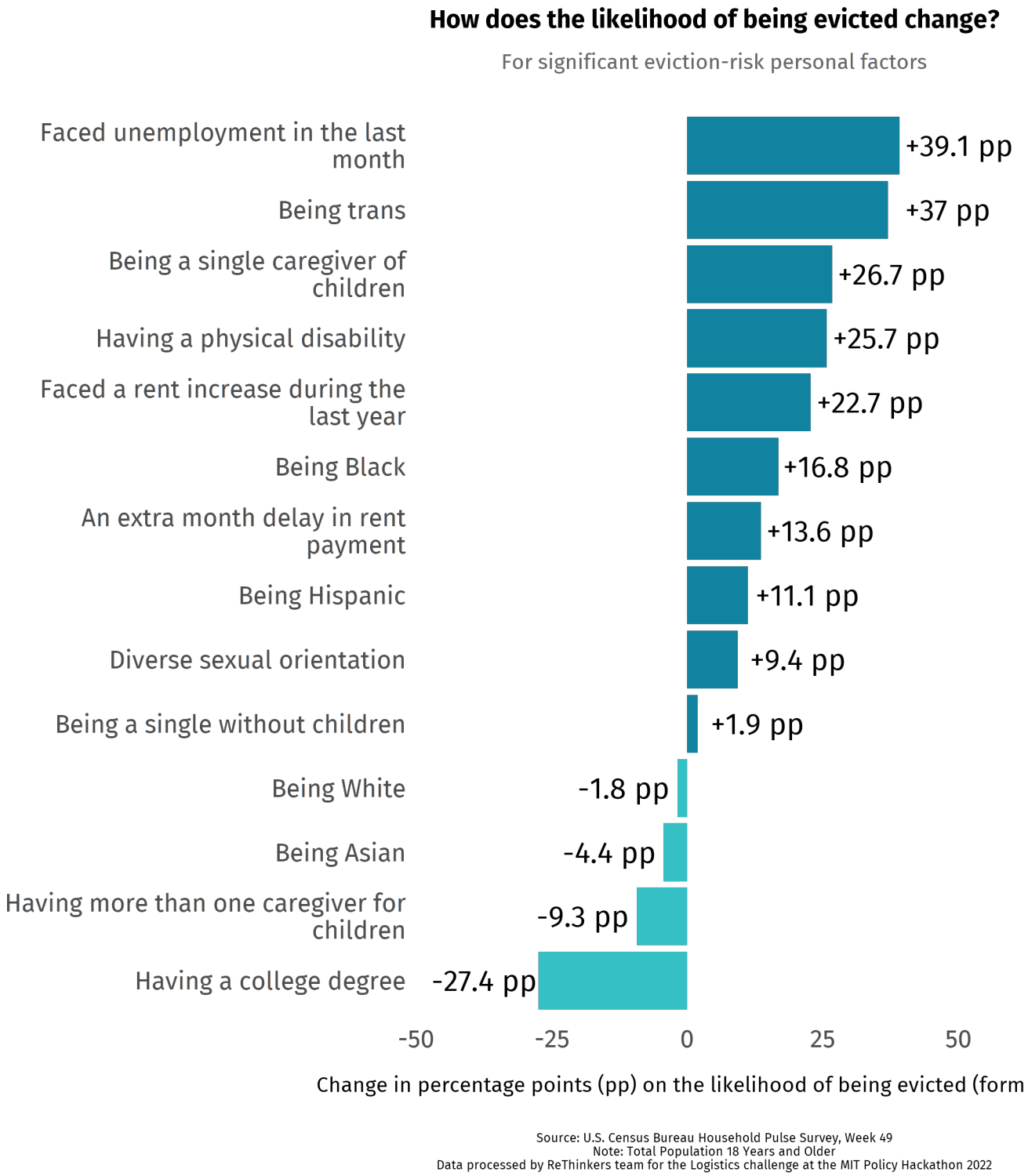
```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

Here's the table of coefficient.

```
##
## Call:
## glm(formula = eviction_likelihood ~ age + female + white + black +
##     asian + hispanic + non_het + trans + single + kids + college +
##     unemployment_month + military + disability_phys + disability_mental +
##     rent_assistance + rent_increase + rent_delay_months + (single *
##     kids), family = binomial(link = "probit"), data = df_clean,
##     weights = df_clean$pweight)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -331.53   -35.84    -6.09    35.47   290.99
##
## Coefficients:
##                       Estimate Std. Error  z value Pr(>|z|)
## (Intercept)        -5.9060104  4.2964095   -1.375   0.1692
## age                -0.0001489  0.0000469   -3.175   0.0015 **
## female             -0.1828730  0.0011886 -153.853  < 2e-16 ***
## white              -0.0434406  0.0021189  -20.502  < 2e-16 ***
## black               0.1434287  0.0021817   65.741  < 2e-16 ***
## asian              -0.0194940  0.0030307   -6.432 1.26e-10 ***
## hispanic            0.1314997  0.0013926   94.428  < 2e-16 ***
## non_het             0.1084790  0.0015415   70.374  < 2e-16 ***
## trans               0.3320077  0.0055782   59.519  < 2e-16 ***
## single              0.0037670  0.0016866    2.233   0.0255 *
## kids               -0.0802096  0.0019753  -40.607  < 2e-16 ***
## college            -0.2647078  0.0013700 -193.213  < 2e-16 ***
## unemployment_month  0.3799888  0.0011967  317.528  < 2e-16 ***
## military            5.0417711  4.2964074    1.173   0.2406
## disability_phys     0.1881264  0.0012715  147.958  < 2e-16 ***
## disability_mental   0.1146980  0.0012116   94.669  < 2e-16 ***
## rent_assistance     0.1407315  0.0016351   86.071  < 2e-16 ***
## rent_increase       0.2144263  0.0011359  188.775  < 2e-16 ***
## rent_delay_months   0.1365571  0.0002486  549.245  < 2e-16 ***
## single:kids         0.2613878  0.0024008  108.878  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4521035  on 860  degrees of freedom
## Residual deviance: 3430427  on 841  degrees of freedom
##   (50076 observations deleted due to missingness)
## AIC: 3434510
##
## Number of Fisher Scoring iterations: 6
```

And a more intuitive way of understanding significant results.



Source: U.S. Census Bureau Household Pulse Survey, Week 49
Note: Total Population 18 Years and Older
Data processed by ReThinkers team for the Logistics challenge at the MIT Policy Hackathon 2022