

PS531 Final

James Baugh

Spring 2021

Questions 1 and 2

Underbanking is the phenomenon of individuals using alternative financial services (AFS) such as payday lenders or check-cashing services rather than using traditional, federally-insured banking services (Rhine and Green, 2012; Burhouse et al, 2016). While research has shown that this phenomenon is incredibly harmful to the personal economic health of those affected (Office of the Inspector General, 2016), and that it disproportionately impacts socioeconomically vulnerable populations in the U.S. (King and Saldarriaga, 2017), not enough work has been done to explore the interaction between underbanking and immigration (Pauwels, 2011, Stookey, 2010). Specifically, it has been identified that underbanking work which examines the country of origin of immigrants is necessary, given the vast diversity of backgrounds among U.S. immigrants (Chatterjee and Zahirovic-Herbert, 2012).

Segmented Assimilation, a theoretical perspective on assimilation first detailed by Portes and Zhao (1993), argues that the process of integrating into a new society and the altering of one's behavior to more closely resemble the behavior of the native population is not the smooth "merging of lines" offered by more traditional approaches to assimilation (Brown and Bean, 2006). This study would seek to explore underbanking rates among the U.S. immigrant population from this perspective, by demonstrating that different groups of immigrants are able to integrate into the traditional American banking system at different rates, and therefore presumably face different barriers.

(H1) First-generation immigrants from developing countries are significantly more likely to be underbanked when compared to citizens

(H2) Second-generation immigrants, regardless of country of origin, experience statistically similar underbanking rates when compared to citizens

Segmented assimilation theory suggests that barriers to assimilation exist which can slow down or even stop the process of assimilating into the mainstream behaviors of a native population (Brown and Bean, 2006; Lee, 2009). Banking is a vital but often overlooked segment of these behaviors, which has extensive financial implications for immigrants (Osili and Paulson, 2004; Paulson and Rhine, 2007; Chatterjee and Zahirovic-Herbert, 2012). Given that prior research into banking rates suggests that financial literacy closely resembles the banking rates of one's country of origin (Klapper, Lusardi, and Van Oudheusden, 2015), it is important to try and fill the gap created in extant literature, which has an unfortunate tendency to gloss over huge differences in the U.S. immigrant population (Waters and Jimenez, 2005).

I would expect, given prior research and the suppositions put forward by the theory, to find that immigrants from developing countries, which would have the least developed banking systems, would be more like to experience underbanking. Furthermore, evidence has shown that the children of first-generation immigrants, often referred to as second-generation immigrants, have economic behaviors that much more closely resemble that of citizens, indicating they may face fewer, or at least different, barriers to assimilation (Pew Research Center, 2013).

Questions 3 and 4

This current project is a replication study of a 2016 FDIC report concerning national underbanking rates (Burhouse et al, 2016). That study was based upon analysis of a nationally-representative, randomly sampled survey conducted by the Census Bureau as part of its June 2015 Current Population Survey. An addendum of questions concerning banking behaviors was included in that survey. While over 70,000 households responded to the broader survey, only about half (about 36,000) responded to the underbanking supplement.

I am accepting the framework imposed by the FDIC study as an observational, survey-driven study, so that the analysis performed here can be more directly compared to the original findings of their 2016 report.

The advantages of this research design are plenty: the CPS is a long-running survey being conducted by an agency with the capacity to gather nationally representative data with a random sampling scheme that includes responses in the tens of thousands. This makes inference to the American population easy to justify. Further, the CPS is primarily a demographic and labor survey, which means it includes a robust series of questions concerning socioeconomic status, employment status, and much more. It is this sweeping array of questions that makes decomposing immigrant status for this study possible.

However, there are also disadvantages. There is no logical treatment or experiment variable or structure present in this design, which makes analysis centered on that type of approach seemingly inappropriate. An exceptionally large N is computationally taxing, and indeed as will be discussed later, may even prevent certain calculative functions from successfully running as a result of the size of the output being too large. Further, the study's more than 1200 variables make missing values ubiquitous, and thus, a strategy for dealing with missing values without compromising the integrity of the analysis is necessary.

Question 5

The dependent variable for this analysis, being underbanked, is a binary variable, where a response of 1 indicates that the respondent is underbanked. The FDIC defines underbanking as not having a traditional checking or savings account, or having used a payday loan, a check-cashing service, a pawn shop loan, an auto title loan, a tax refund anticipation loan, a rent-to-own service, purchasing a money order, or sending an international remittance, outside of a traditional bank within the last 12 months. Immigration status in this analysis is being decomposed from the original study's simple binary indicator of being a "foreign-born non-citizen" to being a citizen, a first-generation immigrant from a developing country or developed country, or a second-generation from a developing or developed country. This was done by comparing the place of birth, father's place of birth, and mother's place of birth responses from the broader CPS survey to the United Nation's 2014 Country Classification Index to determine the development status of a country of origin. Further variables included in the original study and replicated here: race (white, black, Hispanic, Asian, other), education (less than high school, high school, some college, college degree or more), annual income (a continuous variable centered to the mean), employment status (employed, unemployed, not in labor force), age, sex, marital status (married, never married, widowed/separated/divorced), and number of children.

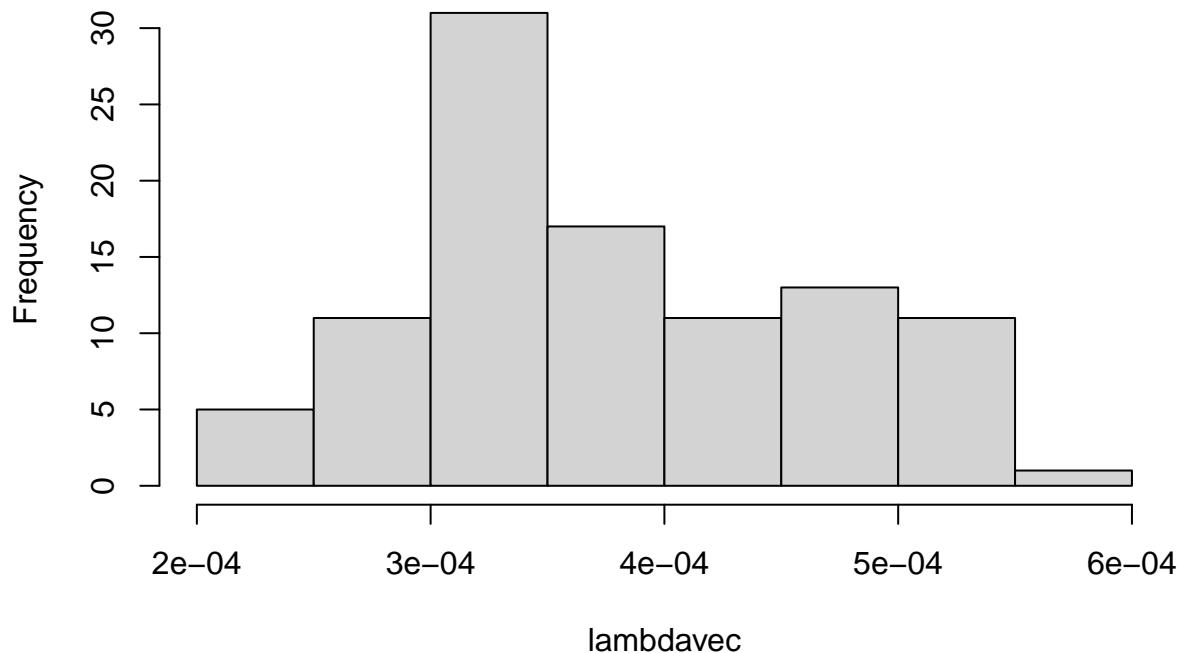
Question 6 A and B: Research Design

Given the large amount of proposed covariates and inherent interconnection between the many facets of socioeconomic status, it is necessary to ensure that this analysis includes these covariates and adjusts for them knowing the potential affect this may have on what the resulting coefficients mean. This study will be employing a penalization strategy for regularizing covariates to allow for better statistical adjustment. Specifically, it will be employing a LASSO-driven strategy to remove from the model any covariates found to have an unsubstantial correlation with the dependent variable. This allows for a strategy of model selection which attempts to both capture the benefits of including covariates (in this case, the examination of the relationship between underbanking and immigration status absent the effect of the many covariates) while minimizing the problems which belabor the interpretability of multiple regression results through model simplification (James et al, 2013).

Traditional LASSO is not a panacea, however. LASSO is agnostic when it comes to removing covariates that are part of a categorical response variable, complicating the position of the remaining categories in the

model. Group LASSO is an alternative method to the traditional LASSO regularization which attempts to correct for this issue, however, as will be seen, this additional adjustment was not necessary in this analysis. Furthermore, the resulting coefficients in a LASSO model can be difficult to interpret, given the penalty imposed on them was calculated based on covariates that are removed from the final model. Therefore, this adjustment will be performed using the Relaxed LASSO (Meinshausen, 2007; Hastie, Tibshirani, and Friedman, 2009). In short, this process entails first performing the traditional LASSO, and then repeating the process of calculating the penalty for the model after removing all covariates with effects reduced to zero from the first model, and applying this new penalty to your simplified model. This process can be repeated until no covariates are reduced to zero in the newest model, as is done here. In theory, and as shown by experimental design and testing, the Relaxed LASSO's penalty is more accurately calculated to those covariates determined by the LASSO process to have non-zero effects on the dependent variable, while still capturing the benefits of model simplification (Hastie, Tibshirani, and Tibshirani, 2017). Additionally, given that the process of calculating the penalty is based upon randomized k-fold cross validation, which can be subject to high variance, it is important to ensure that the penalty is calculated conservatively (Hastie, Tibshirani, and Friedman); this is accomplished here by calculating the penalty 100 times in each LASSO iteration and taking the mean of these penalties as the penalty to be applied.

Histogram of Calculated Penalties



##	Covariates	Estimates
##	[1,] "First-generation, developing country"	"0.12647"
##	[2,] "First-generation, developed country"	"0"
##	[3,] "Second-generation, developing country"	"-0.01967"
##	[4,] "Second-generation, developed country"	"-0.05151"
##	[5,] "Black"	"0.24423"
##	[6,] "Hispanic"	"0.12632"
##	[7,] "Asian"	"-0.01582"
##	[8,] "Other"	"0.15128"
##	[9,] "Less Than High School"	"0.19737"

```

## [10,] "High School"          "0.09345"
## [11,] "Some College"        "0.0594"
## [12,] "Annual Income"      "-0.00098"
## [13,] "Unemployed"         "0.05734"
## [14,] "Not in Labor Force"  "0.01411"
## [15,] "Age"                 "-0.00359"
## [16,] "Sex"                 "0"
## [17,] "Married"             "-0.05377"
## [18,] "W/S/D"               "-0.01526"
## [19,] "Num. of Children"    "0.01546"

##      Penalty 1 Penalty 2
## [1,] 0.00038 0.00024

##      Covariates      Estimates
## [1,] "First-generation, developing country" "0.12703"
## [2,] "Second-generation, developing country" "-0.02087"
## [3,] "Second-generation, developed country" "-0.05268"
## [4,] "Black" "0.24443"
## [5,] "Hispanic" "0.12648"
## [6,] "Asian" "-0.01662"
## [7,] "Other" "0.15196"
## [8,] "Less Than High School" "0.19809"
## [9,] "High School" "0.09419"
## [10,] "Some College" "0.06014"
## [11,] "Annual Income" "-0.00098"
## [12,] "Unemployed" "0.05775"
## [13,] "Not in Labor Force" "0.01494"
## [14,] "Age" "-0.0036"
## [15,] "Married" "-0.05456"
## [16,] "W/S/D" "-0.0161"
## [17,] "Num. of Children" "0.01563"

```

Given that one of the central purposes of LASSO-driven covariance adjustment is model simplification and selection, an important metric of if this process is successful in this analysis is whether the model becomes simpler, IE are covariates selected for removal. If none are, or if only a few are, it would indicate that LASSO is not able to substantially simplify the model from its currently complex state. A secondary test for Relaxed LASSO is whether the iterative penalties are smaller values than the prior penalty – a smaller value is to be expected and would indicate less “noise” when it comes to examining the relationship between the dependent variable and the covariates (Hastie, Tibshirani, and Friedman, 2009). As can be seen above, while the second calculated penalty is indeed smaller than the first, only two covariates had their estimates reduced to zero, and no additional covariates were removed in the second iteration of the LASSO, meaning that the model simplification is minor in this analysis. This is difficult to interpret, and may mean that LASSO as a strategy of regularization doesn’t work quite that well when applied to this particular case.

Question 7

As stated before, missing values are ubiquitous in this data, given its length and massive pool of respondents. Because LASSO cannot function properly with missing values in the dependent variable, observations with missing values in the underbanking indicator are removed from the dataset used to perform this analysis. While more than 1,000 observations are removed, the remaining size of N is still quite vast – 34,243.

Extreme outcome or covariate values are a bit more difficult to assess. Given that underbanking is used here as a binary variable, extreme outcomes would be presumably impossible. However, there may be extreme covariate values (extremely poor or wealthy respondents, extremely old or young respondents, etc.). Yet it is unlikely that these cases would have an undue influence on the results of the analysis, given the large size of

N. Additionally, these extreme values are worth direct attention: if the extremely poor are much more likely to be underbanked, which is likely given prior literature and research, that should be examined directly, not adjusted for. Importantly, it does not seem to be the case that extreme values were adjusted for in any way in the FDIC’s analysis, and thus, it does not seem appropriate to do so here either.

Question 8, 9, and 10: Statistical Tests

This analysis is largely focused on population inference, and as such, it will perform p-value tests of significance through an OLS regression fit on the simplified model defined by the Relaxed LASSO. Given that LASSO directly performs penalization on the coefficients it generates, it is an extremely biased and difficult to interpret approach (Liu, Sun, and McGovern, 2017). While much work is being done on the performance of significance tests on LASSO models (Lockhart et al, 2014; Lee et al, 2016; Taylor and Tribshirani, 2018), it was not able to be successfully performed for this analysis. The ‘selectiveInference’ package (see <https://cran.r-project.org/web/packages/selectiveInference/selectiveInference.pdf>) was explored, and the coding has been included in the appendix, but error codes were generated related to the size of the output which were unable to be resolved. As a result, the model selected by Relaxed LASSO was used to perform a linear regression with HC2 robust standard errors to create the p-values used to perform the significance tests. This method, while less desirable than direct inferential calculation from the Relaxed LASSO, has been nonetheless argued to be a valid way to correct for the biases inherent to LASSO when it comes to making interpretations and inferences (Hastie, Tibshirani, and Friedman, 2009).

Given the simple and singular nature of the test, I decided to judge the performance of the test by evaluating its false positive rate, its power, and its p-adjusted false discovery rate. A family-wise error rate seemed redundant or potentially even inappropriate given that only one test is being performed which can already easily be judged along these other metrics. False discovery rate was selected because it should ensure that less than 5% of significant tests will result in false positives, rather than the 5% of all tests measured by simple false positive rates, generating a more conservative estimate (Non-linear Dynamics, n.d.).

Power was determined to be ‘1’, indicating that the test has a 100% chance to detect the hypothesized effect, given this data’s N and the model’s degrees of freedom. This seems intuitively inaccurate, but was calculated using the ‘pwr’ package, which was employed correctly according to its own documentation as well as according to a secondary source (<https://data-se.netlify.app/2018/07/24/power-calculation-for-the-general-linear-model/>). This, if correct, would indicate the test performs perfectly.

False positive rate and false discovery rate were calculated by, in the case of the former, examining the p-values generated by the linear regression function in R, and in the case of the latter, using the ‘p.adjust’ function (see <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/p.adjust.html>) in R set to the ‘BH’ method for false discovery rate calculation. The below table shows the two rates side-by-side for easy comparison. They are identical in most cases, which is likely the result of the large N ensuring high confidence in the selected testing metrics. The results indicate that for most covariates, we can be confident of their ability to estimate the national rate of underbanking, with the risk of false positivity being less than 0.1% in all but 4 instances.

```
##
##      Multiple regression power calculation
##
##              u = 17
##              v = 34225
##              f2 = 0.189485
##      sig.level = 0.05
##              power = 1
##
##              FPR  FDR
## (Intercept)    0.00 0.00
## fgen_developing 0.00 0.00
## sgen_developing 0.01 0.01
## sgen_developed  0.21 0.22
```

```
## black          0.00 0.00
## hispanic1      0.00 0.00
## asian          0.24 0.24
## other1         0.00 0.00
## lessthanhs     0.00 0.00
## hs             0.00 0.00
## somecollege    0.00 0.00
## famincmean     0.00 0.00
## unemployed     0.00 0.00
## nilf           0.02 0.02
## age            0.00 0.00
## married        0.00 0.00
## wsd_married    0.02 0.02
## numchildren2   0.00 0.00
```

Questions 11, 12, and 13: Statistical Estimation

The estimators used in the study are the coefficients (the calculated beta values) from the linear regression model, and they are estimating the underbanking rates of the U.S. population among those same groups (in other words, the population rates are the estimand). This was chosen because they are more reliable than the LASSO coefficients, as explained above. Further, even though the FDIC study employed logistic regression, linear regression results with binary dependent variables are reliable and more intuitive than the odds ratios output by logit modeling (Hellevik, 2009). As Hellevik explains, the coefficients shown below are not odds ratios as they would be in a logistic regression. Instead, they are proportional differences. Interpreting them is straight-forward: a coefficient of approximately 0.128 for first-generation immigrants from developing countries means that the proportion of those in this group who are underbanked is 0.128 higher than the proportion of the underbanked among the group to which they are being compared, in this case, citizens and first-generation immigrants from developed countries.

```
# Apologies for the raw variable names
cbind(Estimates=round(coef(linreg), 5))
```

```
##              Estimates
## (Intercept)    0.41707
## fgen_developing 0.12790
## sgen_developing -0.05476
## sgen_developed -0.02297
## black          0.24472
## hispanic1      0.12672
## asian          -0.01790
## other1         0.15312
## lessthanhs     0.19959
## hs             0.09564
## somecollege    0.06152
## famincmean     -0.00097
## unemployed     0.05842
## nilf           0.01622
## age            -0.00361
## married        -0.05605
## wsd_married    -0.01766
## numchildren2   0.01594
```

Bias, mean square error (MSE) and root mean square error (RMSE) will be examined to judge the validity of these estimators. Bias is to be expected, given the strategy of model selection, as was discussed above. However, this bias is likely minimal given the small number of variables reduced to zero by LASSO, and this

increase in bias should be worth the reduction in variance offered by the (albeit slightly) simplified model. RMSE was chosen to compliment MSE due to its ease of interpretability, as it is output in the same unit as the dependent variable.

Bias was calculated manually by subtracting the mean value of the observed cases of the dependent variable from the fitted values of the dependent variable provided by the regression model (Dalpiaz, 2020). The results, shown below, indicate that bias is functionally 0, indicating that the estimators do not systematically over- or under-estimate the population parameter, in this case, national underbanking rates.

MSE was also calculated manually by taking the squared mean of the difference between the observed values of the dependent variable and the fitted values of the regression model, and the square root of that value for RMSE (Dalpiaz, 2020). Given that only one set of estimators was calculated for this analysis, the traditional evaluation of MSE-based performance for a model (IE, comparing it to other models to see which has the lowest MSE) is not available. Additionally, because of the binary response categories of the dependent variable, it is conceptually difficult to interpret RMSE as well (0.407 of being underbanked?). If bias is truly zero, then MSE and RMSE are both comprised entirely of the variance in the estimators, which would be expected to be high given “all or nothing” nature of binary variables. I am comfortable with this level of variance, as it seems unlikely that any model would be able to substantially reduce this variance, and doing so would likely come at the cost of increasing bias regardless.

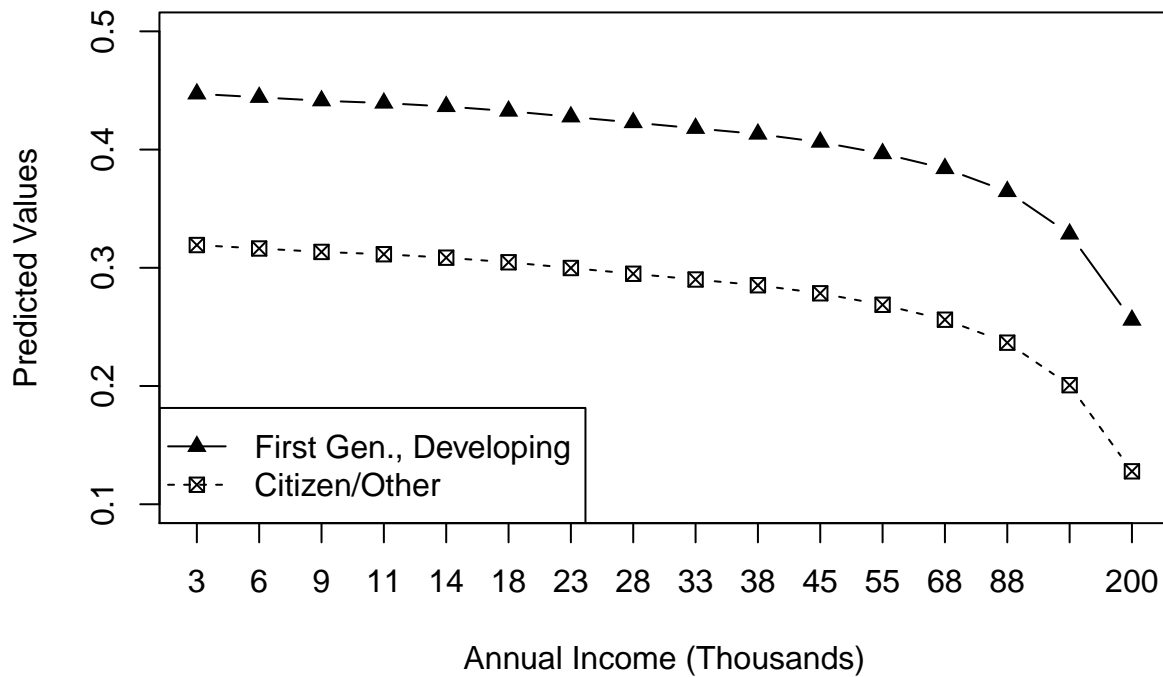
```
##           MSE           RMSE
## [1,] 0.1657715 0.4071504

##           [,1]
## Mean Fitted    2.701866e-01
## Mean Observed  2.701866e-01
## Bias           -1.931788e-14
```

Question 14

The figure designed for this project, shown below, is a graph of predicted values of underbanking by income, between first-generation immigrants from developing countries and everyone else (citizens and immigrants of other categories). As can be seen, the values (which can be interpreted as the proportional difference of underbanking rates) of the first-generation immigrants from developing countries are substantially higher than their counterparts at all levels of income. Remarkably, the value for this group at the highest level of income is approximately equal to the value of their counterparts at the lowest level of income, indicating that first-generation immigrants from developing countries experience substantially higher rates of underbanking across the income spectrum to a severe degree.

Figure: Predicted Underbanking Values by Income



Appendix A: Repository Link

Here is a link to the github repository. I'm still not super familiar with how this works, so please reach out to me (baugh2@illinois.edu) if this doesn't work correctly for whomever is reading this. You should find the r script, the compiled pdf, and the Rmd file used to create it here. Unfortunately, Github is saying that the dataset is too large (>25mb) to be uploaded there, but it is of course available upon request.

The link: PS531 Stats Final Repository Spring 2021

Appendix B: References

Brown, S., & Bean, F. (2006). Assimilation models, old and new: Explaining a long-term process. Migration Policy Institute. Retrieved from: <https://www.migrationpolicy.org/article/assimilation-models-old-and-newexplaining-long-term-process>

Burhouse, S., Chu, K., Ernst, K., Goodstein, R., Lloro, A., Lyons, G., . . . Weinstein, J. (2016). FDIC national survey of unbanked and underbanked households. Retrieved from <https://www.fdic.gov/household-survey/2015/2015report.pdf>

Chatterjee, S., & Zahirovic-Herbert, V. (2012). A road to assimilation: Immigrants and financial markets. *Journal of Economics and Finance*, 38, 345-358. Retrieved from: <https://link.springer.com/article/10.1007/s12197-011-9224-5>

Dalpiaz, D. (2020). R for Statistical Learning. Retrieved from: <https://daviddalpiaz.github.io/r4sl/index.html>

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer Science & Business Media.

- Hastie, T., Tibshirani, R., and Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. arXiv preprint arXiv:1707.08692. Retrieved from: <https://arxiv.org/pdf/1707.08692>.
- Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Qual Quant*, 43, 59-74. DOI: 10.1007/s11135-007-9077-3
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
- King, D., & Saldarriaga, J. (2017). Access to taxicabs for unbanked households: An exploratory analysis in New York City. *Journal of Public Transportation*, 20(1), 1-19. <https://doi.org/10.5038/2375-0901.20.1.1>
- Klapper, L., Lusardi, A., & Van Oudheusden, P. (2015). Financial literacy around the world: Insights from the Standard & Poor's rating services global financial literacy survey. Global Financial Literacy Excellence Center. Retrieved from: https://responsiblefinanceforum.org/wp-content/uploads/2015/12/2015-Finlit_paper_17_F3_SINGLES.pdf
- Lee, C. (2009). Sociological theories of immigration: Pathways to integration for U.S. immigrants. *Journal of Human Behavior in the Social Environment*, 19(6), 730- 744. Retrieved from: <https://www.tandfonline.com/doi/abs/10.1080/10911350902910906>
- Lee, J., Sun, D., Sun, Y., and Taylor, J. (2016). "Exact post-selection inference, with application to the lasso." *Ann. Statist.* 44(3), 907-927. <https://doi.org/10.1214/15-AOS1371>
- Liu, Z., Sun, F. & McGovern, D.P. Sparse generalized linear model with L 0 approximation for feature selection and prediction with big omics data. *BioData Mining* 10, 39 (2017). <https://doi.org/10.1186/s13040-017-0159-z>
- Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the lasso. *Annals of statistics*, 42(2), 413. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4285373/>
- Meinshausen, N. (2007). Relaxed Lasso. *Computational Statistics & Data Analysis*, 52(1), 374-393. Retrieved from: <https://www.sciencedirect.com/science/article/abs/pii/S0167947306004956>
- Non-linear Dynamics. (n.d.). P-values, False Discovery Rate (FDR) and q-values. Retrieved from: <https://tinyurl.com/7pwhenj4>
- Office of the Inspector General, United States Post Office. (2014, January 27). Providing non-bank financial services for the underserved. US Post Office of the Inspector General White Paper, Report Number RARC-WP-14-007. Retrieved from https://www.uspsoig.gov/sites/default/files/document-library-files/2015/rarc-wp-14-007_0.pdf
- Ondersma, C. (2016). Debt without relief: An empirical study of undocumented immigrants. *Rutgers University Law Review*, 68, 1801-1840. Retrieved from <http://www.rutgerslawreview.com/wp-content/uploads/2017/05/Chrystin-Ondersma-Debt-Without-Relief-68-Rutgers-U.-L.-Review-1801-2016.pdf>
- Osili, U., & Paulson, A. (2004). Prospects for immigrant-native wealth assimilation: evidence from financial market participation. Federal Reserve Bank of Chicago, Working Paper 2004-18. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=624187
- Paulson, A., & Rhine, S. (2007). The financial assimilation of an immigrant group: Evidence on the use of checking and savings accounts and currency exchanges. *Journal of Family and Economic Issues*, 2(29), 264-278. Retrieved from: <https://link.springer.com/article/10.1007/s10834-008-9097-8>
- Pauwels, M. (2011). Ethnicity and financial exclusion: How fringe banking has taken hold in ethnic and immigrant neighborhoods. *Ethnic Studies Review*, 34(1), 211- 219. Retrieved from <https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1362&context=esr>
- Pew Research Center (2013, February 7). Second-Generation Americans. Retrieved from: <http://www.pewsocialtrends.org/2013/02/07/second-generation-americans/>

- Portes, A., & Zhou, M. (1993). The new second generation: Segmented assimilation and its variants. *The annals of the American academy of political and social science*, 530(1), 74-96.
- Rhine, S., & Greene, W. (2012). Factors that contribute to becoming unbanked. *Journal of Consumer Affairs*, 47(1), 27-45. doi:10.1111/j.1745-6606.2012.01244.x
- Stookey, S. (2010). Financial services segregation: Improving access to financial services for recent Latino immigrants. Inter-American Development Bank. Retrieved from <https://publications.iadb.org/bitstream/handle/11319/6262/Financial%20Services%20Segregation%3A%20Improving%20Access%20to%20Financial%20Services%20for%20Recent%20Latino%20Immigrants.pdf?sequence=1>
- Taylor, J., & Tibshirani, R. (2018). Post-selection inference for-penalized likelihood models. *Canadian Journal of Statistics*, 46(1), 41-61. Retrieved from: <https://arxiv.org/pdf/1602.07358.pdf>
- Waters, M., & Jimenez, T. (2005). Assessing immigrant assimilation: New empirical and theoretical challenges. *Annual Review of Sociology*, 31, 105-125. Retrieved from: <https://dash.harvard.edu/handle/1/3203280>