URBAN INSTITUTE · ELEVATE · THE · DEBATE

Young Adults' Use of Credit and Debt During the COVID-19 Pandemic

Technical Appendix for Data Tables

Kassandra Martinchek

Last Updated January 2024

This technical appendix briefly describes the empirical strategy for the data tables published under the " Young Adults' Use of Debt and Credit During the COVID-19 Pandemic Data Tables" on the Urban Institute Data Catalog (Accessible from https://datacatalog.urban.org/dataset/young-adults-use-debt-and-credit-during-the-pandemic. Data originally sourced from credit bureau data, developed at the Urban Institute, and made available under the ODC-BY 1.0 Attribution License).

Data Details

The core analytic dataset for this project is derived from a 2 percent nationally representative sample of 5 million consumer credit records provided by one of the three major credit bureaus. These data are longitudinal, following the same consumers over time, and are refreshed at each data pull to maintain the sample's representativeness. Consumer credit records contain details on consumers' zip code of residence, age, credit scores, debt amounts, delinquencies, and ownership of various loans and accounts – but do not contain details on consumers' race and ethnicity. Notably, the data do not include details on 11 percent of U.S. adults with no credit record, with people of color and young adults disproportionately represented among credit invisibles (Brevoort et al., 2015).

Variable Definitions

In this study, I use the following variables to capture consumers' credit and debt over the course of the COVID-19 pandemic, following variable definitions established by Martinchek and Braga (2022):

- **Credit scores**: the average VantageScore from 300 to 850, where scores below 600 are considered subprime.
- **Auto and retail loan delinquencies**: an indicator for whether consumers who have an open auto or retail loan are 60 days or more past due on payments on these loans.
- **Credit card utilization**: the average share of available credit card credit used across all open credit cards among consumers with at least one open credit card.
- **Credit card delinquencies**: an indicator for whether consumers who have at least one open credit card are 30 days or more past due on payments on at least one credit card.
- **Alternative financial services loan use**: an indicator for whether consumers with a credit record have an open alternative financial services loan. Alternative financial services loans include short-term unsecured loans (such as payday loans), loans where personal property was used as collateral (such as auto title loans), or transactions under which property was leased in exchange for a weekly or monthly payment with the option to purchase (rent-to-own) from online small-dollar lenders; online installment lenders; storefront small-dollar lenders; and single payment, line of credit, auto title, and rent-to-own lenders.

Sample Characteristics Tables and Figures

The sample characteristics tables include descriptive characteristics of the analytic sample used, including attrition rates over the panel period between February 2020 and August 2023, loan use and credit scores at baseline, the share of the sample that lives in each community, age breakdown of the sample, and credit score distributions of the matched and unmatched cohorts of young adults. Consult the table notes of each table for what summary statistics it contains.

The sample characteristics figures contain figures of the distribution of credit scores at baseline for the matched and unmatched cohorts of young adults ages 20-23, 24-26, and 27-29.

Community Comparison Tables for Young Adults

These tables present regression output for an analysis of trends in young adults' debt and credit between February 2020 and August 2023 across community demographic compositions. Each table presents output for a different credit and debt outcome of interest, including: (1) credit scores, (2) auto and retail loan delinquencies, (3) credit card utilization, (4) credit card delinquencies, and (5) alternative financial services loan use (see variable definitions above). Regression output included in these tables are from an individual-consumer level regression with interacted and main effects for community racial and ethnic composition and time, with no other covariates and errors clustered at the Zip Code Tabulation Area level. Community composition is coded as a binary indictor that equals one if consumers live in a community of color in February 2020 and zero if they live in a majority-white community. The time variable is a categorical indicator that numbers the data extract and ranges from 0 to 13, capturing 13 data extracts between August 2018 and August 2023. There are separate specifications for each community demographic, including majority-Black, majority-Hispanic, and majority-Native American communities (or Zip Code Tabulation Areas where 50 percent or more of residents identify as a particular race or ethnicity in the 2015-2019 American Community Survey). The interacted effect of this regression specification captures *the change in the difference* in credit and debt outcomes between young adults living in communities of color and those living in majority-white communities (relative to what it was in February 2020) and is useful for assessing whether community-level racial disparities in credit and debt widen or narrow over the course of the pandemic (between February 2020 and August 2023).

EQUATION A.1

Estimation Strategy for Community Comparison Analysis

$$Y_{it} = \alpha_1 Community_i + \alpha_2 Community_i \times \theta_t + u_{izt}$$

Here, Y_{it} represents the credit or debt outcome of interest of young adult i in period t, α_1 is a measure of the difference in credit outcomes between young adults living in communities of color in February 2020, relative to majority-white communities , α_2 allows for the time trend in the outcome variable to vary based on whether or not a consumer lives in a community of color in February 2020, θ_t represents time fixed effects, and u_{izt} is the heteroskedastic-robust error term, which is clustered at the Zip Code Tabulation Area level. This specification is run separately for each outcome and for majority-Black, majority-Hispanic, and majority-Native American communities. Throughout the analysis, we characterize whether a consumer lives in a community of color or majority-white community based on their zip code of residence in February 2020.

Subgroup Comparison Tables

These tables present regression output for an analysis of credit score trends for young adults with different credit score profiles, as measured by credit tier (or whether they had prime or subprime credit scores at baseline) and student loan status (or whether they had a student loan in February 2020). Regression output included in these tables are from two analyses: (1) an individual-consumer level regression with interacted and main effects for subgroup membership and time, with no other covariates and errors clustered at the Zip Code Tabulation Area level; and (2) an individual-consumer level regression with interacted and main effects for community racial and ethnic composition and time, with no other covariates and errors clustered at the Zip Code Tabulation Area level. In both, the time variable is a categorical indicator that numbers the data extract and ranges from 0 to 13, capturing 13 data extracts between August 2018 and August 2023.

In analysis #1, subgroup membership is 1 if a consumer has subprime credit or a student loan in February 2020 (separately) and 0 if they do not. The interacted effect of this regression specification captures the change in the difference in credit scores between young adults with either subprime credit or student loans and those who do not have these characteristics and is useful for assessing whether young adults with student loans or subprime credit experience sharper or less pronounced gains in credit health over the course of the pandemic, relative to non-student loan holders and prime consumers.

In analysis #2, community composition is coded as a binary indictor that equals one if consumers live in a community of color in February 2020 and zero if they live in a majority-white community. There are separate specifications for each community demographic, including majority-Black, majority-Hispanic, and majority-Native American communities (or Zip Code Tabulation Areas where 50 percent or more of residents identify as a particular race or ethnicity in the 2015-2019 American Community

Survey) and for subsamples of consumers with subprime credit, prime credit, student loans, and no student loans (for a total of 12 regressions). The interacted effect of this regression specification captures the change in the difference in credit and debt outcomes between young adults living in communities of color and those living in majority-white communities and is useful for assessing whether community-level racial disparities in credit scores widen or narrow over the course of the pandemic (between February 2020 and August 2023). This analysis captures if trends in community-level racial disparities in credit scores vary across subgroups (or consumers with subprime credit, prime credit, student loans, and no student loans at baseline in February 2020).

Oaxaca-Blinder Decomposition Tables

These tables present output from a three-fold Oaxaca-Blinder decomposition that tests if community-level racial disparities in credit scores in February 2020 are due to variations in common correlates of financial well-being (e.g., homeownership, employment, income, and educational attainment) or due to either unobserved or structural factors following methodology described by Rahimi and Hashemi Nazari (2021). I use Zip Code Tabulation Area data from the 2015-2019 American Community Survey to decompose differences in average credit scores of young adults living in majority-Black, majority-Hispanic, and majority-Native American communities and those living in majority-white communities. The term for "coefficients" reflects the portion of the mean difference in credit scores between group 1 (young adults living in majority-white communities) and group 2 (young adults living in communities of color) that are not explained by differences in community-level covariate levels and captures (1) differences in rates of returns to education, homeownership, employment, and income experienced by young adults living in different communities, (2) omitted covariates, and (3) the effect of discrimination and community-level structural racism.

Age Group Comparison Tables

These tables present regression output for an analysis of trends in consumers' debt and credit between February 2020 and August 2023 across age groups. Each table presents output for a different credit and debt outcome of interest, including: (1) credit scores, (2) auto and retail loan delinquencies, (3) credit card utilization, (4) credit card delinquencies, and (5) alternative financial services loan use (see variable definitions above). Regression output included in these tables are from an individual-consumer level regression with interacted and main effects for age group and time, with no other covariates and errors clustered at the Zip Code Tabulation Area level. Age groups are defined as follows: 1 = ages 20-29, 2 = ages 30-39, 3 = ages 40-49, 4 = ages 50-64, and 5 = older than 65. The time variable is a categorical indicator that numbers the data extract and ranges from 0 to 13, capturing 13 data extracts between August 2018 and August 2023. There are separate specifications for the overall sample and sub-samples of consumers living in majority-Black, majority-Hispanic, and majority-Native American communities (or Zip Code Tabulation Areas where 50 percent or more of residents identify as a particular race or ethnicity in the 2015-2019 American Community Survey). The interacted effect of this regression specification captures the change in the difference in credit and debt outcomes between young adults ages 20-29 and each older age group (relative to what it was in February 2020) and is useful for assessing whether young adults saw relative improvements or reductions in credit between February 2020 and August 2023 compared with older adults.

Cohort Comparison Tables

These tables present descriptive summary statistics and regression output for an analysis of three-year trends in young adults' credit scores across two matched cohorts: young adults ages 20-23, 24-26, and 27-29 in 2016 and 2020. Before conducting regression and descriptive analyses, young adults in the 2020 cohort were matched to young adults in the 2016 cohort by age and baseline credit score using exact matching. Unmatched observations and observations in both cohorts are dropped from analyses and all descriptive statistics and regression analyses are conducted on the matched sample and weighted using the weights generated during exact matching.

Descriptive summary tables include details on average credit scores over time for young adults in each cohort both overall and broken down by specific community demographics.

Regression output included in these tables are from two analyses: (1) an individual-consumer level regression with interacted and main effects for cohort and time, with no other covariates and errors clustered at the Zip Code Tabulation Area level; and (2) an individual-consumer level regression with interacted and main effects for community racial and ethnic composition and time period, with no other covariates and errors clustered at the Zip Code Tabulation Area level. In both analyses, the time variable is a categorical indicator for the number of years after the baseline period.

In analysis #1, there are separate specifications for the overall sample and sub-samples of consumers living in majority-Black, majority-Hispanic, and majority-Native American communities (or Zip Code Tabulation Areas where 50 percent or more of residents identify as a particular race or ethnicity in the 2015-2019 American Community Survey). The interacted effect of this regression specification captures the additional change in average credit scores of young adults in the 2020 cohort relative to the 2016 cohort and is useful for assessing whether young adults in the 2020 cohort saw relative improvements or reductions in credit health over a three-year period compared with similar young adults in the 2016 cohort.

In analysis #2, community composition is coded as a binary indictor that equals one if consumers live in a community of color in February 2020 and zero if they live in a majority-white community. There are separate specifications for each community demographic, including majority-Black, majority-Hispanic, and majority-Native American communities (or Zip Code Tabulation Areas where 50 percent or more of residents identify as a particular race or ethnicity in the 2015-2019 American Community Survey) and each cohort (for a total of 6 regressions for young adults ages 20-23, 24-26, and 27-29). The interacted effect of this regression specification captures the change in the difference in credit scores between young adults living in communities of color and those living in majority-white communities and is useful for assessing whether community-level racial disparities in credit scores widen or narrow over the course of the pandemic (between February 2020 and August 2023). This analysis captures if trends in community-level racial disparities in credit scores vary across cohorts.

Policy Impact Estimate Tables

In these tables, I present impact estimates of pandemic-era state-level consumer protection and safety net policies, including: (1) unemployment insurance extended benefits programs for 13 and 20 weeks, and (2) utility shutoff moratoria. The policy impact tables include estimates from the preferred specification (see causal policy impact analysis section below), as well as several robustness checks.

Causal Policy Impact Analysis

I used a consumer-level staggered difference-in-difference research design to identify the effects of state utility shutoff moratoria and extended unemployment insurance (UI) benefits (both 13 and 20 week programs) on young adults' credit and debt outcomes (equation A.2). Using a difference-in-difference model, I compare young adults' mean credit scores and credit card delinquency rates between states that did and did not implement the policy before and after policy implementation. The underlying assumption was that those affected by the policy and the comparison group would have parallel outcome trends in the absence of state consumer protection or safety net policies.

EQUATION A.2

Estimation Strategy for Difference-in-Difference Policy Impact Estimates

$$Y_{icst} = \gamma_t + \delta_c + \beta Policy_{st} + \beta X_{icst} + \epsilon_{icst}$$

Where Y_{icst} is the credit or debt outcome of interest for young adult *i* residing in county c in state *s*, in period *t*. Throughout the analysis, I characterize individual state and county of residence based on the consumer's home address in February 2020 to account for the potential endogeneity of migration decisions as a response to the policy implementation; γ_t includes year-month fixed effects; and δ_c includes county fixed-effects - while in some specifications, I use individual-fixed effects. By using individual fixed-effects, I control the regression for a consumer's credit history, improving the estimates' precision. The model with individual fixed effects is the preferred model. $Policy_{st}$ are indicators for whether the state *s* had the policy active (utility shutoff moratoria or extended benefits UI programs (13 and 20 week)) in period *t*. X_{ist} is a large set of individual, state, and county-level controls. At the individual level, this vector of controls includes age and age squared. At the state level, I include COVID-19 vaccination rate (population 18+), number of COVID-19 cases per capita and the number of COVID-19 deaths per capita, unemployment rate, the share of UI payments out within three weeks, indicators for whether states had closure orders for restaurants, bars, movie theatres, gyms, and childcare centers, indicators for whether states had active suspensions on vehicle repossessions and garnishments, an indicator for whether states had active Pandemic Unemployment Assistance programs, and an indicator for whether states had an active eviction moratorium in each period. Standard errors are clustered at the state level.

This specification was run on several additional samples: (1) young adults living in different communities (e.g., majority-Black, majority-Hispanic, and majority-Native American communities) and (2) young adults without student loans or mortgages, who likely did not benefit from federal-level forbearances on student loans and mortgage repayment.

Robustness Checks

I performed two robustness checks of the estimates described in equation A.2. These robustness checks were valuable because states experienced different policy and economic shocks during the pandemic, which could confound the impacts estimated in equation A.2, despite the inclusion of controls.

First, I ran equation A.2 on a subsample of consumers living in bordering counties within states that implemented utility shutoff moratoria or 13- and 20-week extended benefits programs and their neighboring counties within states that never implemented that policy using data from the 1991 Census Bureau Contiguous County File, following a similar approach as in Andre et al. (2023a, 2023b). I conduct this analysis separately for each state-level policy of interest, based on the relevant set of bordering counties for each. I adjust the county pair list to keep only counties that share a common land border or are separated by a body of water but connected by a bridge or boat. Contiguous counties were more likely to suffer the same health and economic shocks but differed in their policy responses.

Second, I used policy discontinuities at county borders to identify the causal effects of policies (Dube, Lester, Reich 2010; Schmidt, Shore-Sheppard, and Watson 2020). To perform this analysis, I restructured the data so each county was observed once per period per adjacent pair. This restructuring was necessary so that observations could be assigned a vector of county pair–time fixed effects that allowed the adjacent border county to serve as a counterfactual. I tested several different sets of fixed effects: (1) county-level fixed effects only; (2) pair-time fixed effects onl;, and (3) county-level and pair-time fixed effects (the preferred specification for this robustness check).

EQUATION A.3

Estimation Strategy for Contiguous County Policy Impact Estimates

$$Y_{ict} = \beta_1 Policy_c + \beta_2 X_c + \gamma_{pt} + \varepsilon_{ist}$$

Y_{icst} is the outcome of young adult (age 20 to 29) *i*, living in border county *c* in period *t*. $Policy_c$ is an indicator for whether the adult's county c of residence implemented the policy (utility shutoff moratoria or extended benefits UI programs (13 and 20 week)). X_c includes a robust set of controls, including COVID-19 vaccination rate (population 18+), number of COVID-19 cases per capita and the number of COVID-19 deaths per capita, unemployment rate, the share of UI payments out within three weeks, indicators for whether states had closure orders for restaurants, bars, movie theatres, gyms, and childcare centers, indicators for whether states had active suspensions on vehicle repossessions and garnishments, an indicator for whether states had active Pandemic Unemployment Assistance programs, and an indicator for whether states had an active eviction moratorium in each period. γ_{pt} is a pair-specific time effect. Standard errors are clustered at the state *s* level. This specification is run separately for each policy of interest, based on the contiguous county pairs for that policy in each period.

References

Andre, Jennifer, Breno Braga, Kassandra Martinchek, and Signe-Mary McKernan. 2023a. “The Effects of State Consumer Protection Policies During the Pandemic on the Use of Alternative Financial Services Loans.” Available at SSRN: https://www.doi.org/10.2139/ssrn.4345180.

———. 2023b. “The Effects of the State Utility Shutoff Moratoria on Credit Delinquencies during the Covid-19 Pandemic.” Working Paper.

Brevoort, Kenneth P., Philipp Grimm, and Michelle Kambara. 2015. *Data Point: Credit Invisibles.* Washington, DC: Consumer Financial Protection Bureau (CFPB).

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. “Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties.” *The Review of Economics and Statistics* 92 (4): 945–64.

Martinchek, Kassandra., & Braga, Breno. (2022, March 8). Credit Health during the COVID-19 Pandemic: Technical Appendix. *Urban Institute*. https://urban-data-catalog.s3.amazonaws.com/drupal-root-live/2022/03/07/data%20catalog%20technical%20appendix.pdf

Rahimi, Ebrahim., & Hashemi Nazari, Seyed Saeed. 2021. “A detailed explanation and graphical representation of the Blinder-Oaxaca decomposition method with its application in health inequalities.” *Emerging Themes in Epidemiology*, *18*(1), 1-15.

Schmidt, Lucie, Lara D. Shore-Sheppard, and Tara Watson. 2020. "The impact of the ACA Medicaid expansion on disability program applications." *American Journal of Health Economics 6*, no. 4 (2020): 444-476.

Acknowledgments

This research was funded by the Urban Institute. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the author and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

For more information on this project, see Kassandra Martinchek. 2024. Young Adults' Use of Debt and Credit During the COVID-19 Pandemic Data Tables. Accessible from https://datacatalog.urban.org/dataset/young-adults-use-debt-and-credit-during-the-pandemic. Data originally sourced from credit bureau data, developed at the Urban Institute, and made available under the ODC-BY 1.0 Attribution License.

URBAN INSTITUTE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE