INTERNATIONAL MONETARY FUND

Winning the War? New Evidence on the Measurement and the Determinants of Poverty in the United States

Katharina Bergant, Andrea Medici and Anke Weber

WP/22/4

2022
JAN

WORKING PAPER

 WP/22/4

IMF Working Paper
Western Hemisphere Department

Winning the War? New Evidence on the Measurement and the Determinants of Poverty in the United States

Prepared by Katharina Bergant, Andrea Medici and Anke Weber*

Authorized for distribution by Nigel Chalk
January 2022

ABSTRACT: Using micro-data from household expenditure surveys, we document the evolution of consumption poverty in the United States over the last four decades. Employing a price index that appears appropriate for low income households, we show that poverty has not declined materially since the 1980s and even increased for the young. We then analyze which social and economic factors help explain the extent of poverty in the U.S. using probit, tobit, and machine learning techniques. Our results are threefold. First, we identify the poor as more likely to be minorities, without a college education, never married, and living in the Midwest. Second, the importance of some factors, such as race and ethnicity, for determining poverty has declined over the last decades but they remain significant. Third, we find that social and economic factors can only partially capture the likelihood of being poor, pointing to the possibility that random factors ("bad luck") could play a significant role.

JEL Classification Numbers:	C81, E21, I32, I38.
Keywords:	Poverty, Inequality, Consumption, Provision and Effects of Welfare Programs
Author's E-Mail Addresses:	KBergant@imf.org, AMedici@imf.org, AWeber@imf.org

* The authors would like to thank Jorge Alvarez, Nicholas Carroll, Nigel Chalk, Andrew Hodge, Deniz Igan, Christoffer Koch, Li Lin, Davide Malacrino, Rui Mano, Andrew Tiffin, Yannick Timmer and seminar participants at the IMF for their helpful comments. We are grateful to Bruce Meyer and James Sullivan for generously sharing their codes and helpful discussions on the construction of their poverty measures. All remaining errors are our own. The views expressed in IMF Working Papers are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

WORKING PAPERS

Winning the War? New Evidence on the Measurement and the Determinants of Poverty in the United States

Prepared by Katharina Bergant, Andrea Medici and Anke Weber[1]

[1] The authors would like to thank Jorge Alvarez, Nicholas Carroll, Nigel Chalk, Andrew Hodge, Deniz Igan, Christoffer Koch, Li Lin, Davide Malacrino, Rui Mano, Andrew Tiffin, Yannick Timmer and seminar participants at the IMF for their helpful comments. We are grateful to Bruce Meyer and James Sullivan for generously sharing their codes and helpful discussions on the construction of their poverty measures. All remaining errors are our own. The views expressed in IMF Working Papers are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

"Fighting poverty requires knowing its extent and among whom it is most severe"

—Bruce D. Meyer and James X. Sullivan

1 Introduction

What is poverty and why should we care about it? The literature uses different concepts of poverty, although a common definition is that poverty exists when people lack the means to satisfy their basic needs. The latter can be defined as those necessary for survival or more broadly as reflecting the minimum acceptable standard of living in a given society. Adequate access to goods and services is crucial to be healthy, succeed in school, and the labor market. For instance, McLaughlin and Rank (2018) find that the annual aggregate cost of U.S. child poverty amounts to about 5.4 percent of GDP, reflecting the loss of economic productivity, increased health and crime costs, and increased costs as a result of child homelessness and maltreatment. Given the high societal and individual costs, accurately gauging the extent of deprivation and understanding the causes of poverty is crucial, including to efficiently allocate federal, state, and local funds.

There is no consensus on how best to measure poverty. The U.S. official poverty measure compares a family's income to a poverty threshold that is based on three times the cost of a specific food basket in the 1960s indexed by overall inflation for urban consumers. More recent studies have argued that consumption-based measures of poverty—that draw on household expenditure surveys and attach a consumption equivalent to the value of government assistance—provide a more accurate picture of poverty (Meyer and Sullivan (2012a,b, 2013, 2019)). Consumption based measures can (i) capture non-cash benefits; (ii) account for savings; and (iii) be more accurate (since consumption outcomes are typically better measured for lower income households than income). In addition to absolute measures, there are relative poverty indicators, which measure the share of the population with income or consumption below some percentage of the median.

For absolute poverty measures, the choice of consumer price index (CPI) has important implications for the level of poverty. Work by Meyer and Sullivan (2012a,b, 2013, 2019) indexes poverty thresholds at a rate that is lower than CPI to account for various biases (such as substitution and outlet bias). Based on these assumptions, and unlike official income-based measures, Meyer and Sullivan's consumption-based poverty measure suggests that poverty has fallen dramatically over time. Much of this decline is due to the downward adjustment of inflation, which is at odds with recent studies that have found those at the lower end of the income distribution to face higher, not lower, inflation than measured by the CPI (Jaravel, 2019). There are two potential underlying drivers of such "inflation inequality": different spending patterns across households (with the poor spending disproportionately more on items that have seen higher price increases, e.g., rent) and differences in price increases across goods and services. Jaravel (2019) shows that because companies are increasingly interested in competing for purchases of wealthy individuals, prices for goods that high income households buy are actually decreasing relative to the prices of goods that lower-income families purchase. Intuitively, given that the share of U.S. national income accruing to high-income consumers has steadily increased, firms respond to changes in relative market size by skewing product introductions toward market segments that are growing faster. This process can lead to a decrease in the price of existing products in the fast-growing market segments because increased competitive pressure from new products pushes markups down.

Besides the question of how best to measure poverty, there is the related issue of who are the poor and what determines poverty. Studies have found poverty to be strongly correlated with employment status, education level, marital status, demographic characteristics, and family structure (Hoynes et al., 2006). Recent research shows that poverty is very common among Americans. Rank et al. (2021) find that between the ages of 20 and 75 years, nearly 60 percent of Americans will experience living for at least one year below the official poverty line, while 75 percent of Americans will encounter poverty or near-poverty. The authors attribute this to structural changes in the labor market towards lower-paying jobs coupled with a lack of universal coverage for child care, health care, and affordable housing, which are leaving an increasing number of families economically vulnerable.

This paper focuses on absolute poverty, constructs a consumption-based poverty indicator that considers inflation inequality and then analyzes the determinants of poverty. We first explain how we derive our preferred consumption-based absolute poverty measure and how it compares to other measures. This is followed by a natural experiment of how the expanded child tax credit in the 2021 American Rescue Plan would impact poverty in the U.S. We then turn to the determinants of poverty, using a variety of econometric specifications, including probit, tobit and machine learning models. Machine learning, unlike the more traditional methods of probit or tobit, offers a nonparametric and flexible way to model poverty status. By using machine learning, we hope to develop a superior poverty-identification technique that not only results in higher predictive power but also reveals more clearly the underlying empirical relationships considered. To the best of our knowledge this paper is the first to apply machine learning techniques to examine the determinants of poverty in the U.S.

We find that:

- Relying on a price index that is linked to the consumption patterns of the bottom quintile would mean that the share of the population living in poverty has not fallen materially since 1980. 2019 poverty rates are close to those in the 1980s, although there were up and downward movements over the period, notably during the Global Financial Crisis (GFC) when poverty increased substantially. Child poverty even increased over the last four decades according to this measure. Poverty gaps (the distance of poor households' consumption to the threshold) have increased in nominal terms and only somewhat decreased in real terms over the last four decades.

- These findings are in line with the evolution of other measures of deprivation (such as food insecurity) but contrast with findings of previous studies that compute consumption-based poverty and adjust the overall CPI significantly downwards. Relative poverty has also increased in the U.S. reflecting a more polarized income and wealth distribution.

- The Biden administration's expanded child tax credit would significantly decrease child poverty, by about one-third with our preferred poverty measure that incorporates inflation inequality.

- Families with more children, African American Families, and Hispanic families are more likely to be poor. On the other hand, families where the household head holds a Bachelor's degree or more or where the adults are married are less likely to be poor. Controlling for time dynamics, regional differences have widened since the Global Financial Crisis (GFC), with families located in the mid-west experiencing a higher likelihood of poverty. Moreover, a Bachelor's degree became even more important in the last decade compared to the 1980s while marital status and race became less important in explaining poverty. However, race and ethnicity are still significant determinants of poverty, even after we control for many other characteristics, such as education and employment status. Machine learning models confirm that we are focusing on the most important explanatory variables.

- Our household characteristics are all significant, and our preferred machine learning model specification—a balanced random forest—is able to predict who is poor with a 75 percent accuracy. However, the significant prediction error, of about 25 percent, means that a random component —"bad luck"—could play a significant role in explaining poverty.

The remainder of this paper proceeds as follows. Section 2 summarizes the literature and methodologies on poverty measurement, introduces the data, and shows how U.S. poverty has evolved over the last 40 years. Section 3 presents new facts on poverty determinants based on probit, tobit and machine learning models. Section 4 concludes.

2 Poverty Measurement

2.1 Background

The U.S. Census Bureau releases annual poverty rates based on absolute income thresholds that are adjusted for inflation. The most basic indicator, the official poverty measure (OPM), which is based on household data from the Current Population Survey (CPS), has been published since the 1960s. The OPM compares a family's gross pre-tax money income to a threshold that equates to three times the cost of a minimum food diet in 1963—because 1955 survey data on expenditures suggested that the average family of three or more people allocated about a third of their after-tax income for food. The threshold is adjusted annually by the consumer price index for all urban consumers (CPI-U) (Fisher, 1992). In 2019, the poverty threshold for a two-parent, two-child family was US$25,926. In 2011, the supplemental poverty measure (SPM) was released that accounts for non-cash income public benefits (such as the supplemental nutrition assistance program), while subtracting several categories of expenses from income (tax liabilities, payments for child support, child care and other work expenses, and out-of-pocket medical expenses). To arrive at a threshold, it computes a 5-year moving average of consumer units spending on food, clothing, shelter and utilities adjusted by family size. In 2019, the SPM threshold for a two-adult two-child family amounted to $29,234.

A large literature points to serious flaws in these absolute income-based poverty measures. Most notably, Meyer and Sullivan (2012a,b, 2019) argue that consumption better reflects the material circumstances of disadvantaged families and therefore use expenditure surveys to construct a consumption-based poverty measure. They offer several reasons for why consumption is preferable to income. First, income does not capture consumption smoothing: people can save when income is temporarily high and borrow when it is temporarily low. A retired couple living off their savings may be living quite comfortably even if they have no income. Second, consumption measures will reflect the loss in wealth if asset values fall. It will also reflect consumption out of durables such as housing and cars. Third, consumption is more likely than income to be affected by access to public insurance programs. Thus, it will do a better job of capturing the effects of changes in the government safety net. All of the above may explain why economic models that examine changes in utility use consumption and not income. Finally, there has been a considerable amount of research showing that income is more prone to measurement issues than consumption, especially among the lower income groups that tend to under-report how much they get through independent and/or informal part-time jobs and government assistance programs (Meyer et al., 2009).

To adjust poverty thresholds over time, studies typically use a measure of inflation for urban consumers, as the majority (90 percent) of households represented in consumer surveys live in urban areas. Meyer and Sullivan (Meyer and Sullivan (2012a,b, 2019)) adopt an adjustment to inflation that slows the growth of the poverty threshold over time to reflect various biases in the measurement of the CPI-U. The literature has typically emphasized four types of biases in the CPI-U, namely (i) substitution bias which occurs when the index uses a fixed market basket and consumers substitute away from high-relative-price items, (ii) outlet bias referring to inadequate accounting for the movement of purchases toward low-price discount or big-box stores,

(iii) quality bias resulting from inadequate adjustments for quality improvements in products over time, and (iv) new product bias referring to the omission or long delay in the incorporation of new products into the index. The retroactive series using current methods (CPI-U-RS) seeks to correct for those biases, but some studies conclude that even the CPI-U-RS is biased upwards by about 0.8 percentage points per year (Berndt, 2006; Gordon, 2006).

The recent literature casts doubt on the appropriateness of using a single rate of change in the average price of goods and services faced by the population as a whole. Intuitively, it may well be the case that prices are rising faster for goods and services that dominate consumption by the poor, such as for example rent. Jaravel (2019) formalizes this idea using several price and expenditure datasets, including scanner data collected in retail stores and the consumer expenditure survey, and shows that the annual increase of the CPI-U is 0.4 percentage points higher on average for the bottom income quintile compared with the top income quintile between 2004 and 2015 (results from the scanner data point to a 0.8 percentage points difference between the top and bottom income quintiles). Similar results are obtained with scanner data in other studies (Kaplan and Schulhofer-Wohl, 2017; Argente and Lee, 2021).

2.2 Data and Methodology

The consumption data come from the Consumer Expenditure (CE) Survey, which is the most comprehensive source of consumption data in the U.S. We use the CE Interview Survey component for 1980-81 and 1984-2019. Data are not available for 1982 and there were relatively fewer observations in 1983. From 1980 to 2010 the survey uses a rotating panel that includes about 5,000 families each quarter and about 7,500 families thereafter. Each consumer unit reports spending on a large number of expenditure categories for up to four consecutive quarters. The survey also has information on consumer unit income, and demographic characteristics. The data are a repeated cross section as the survey does not follow the same units over time.

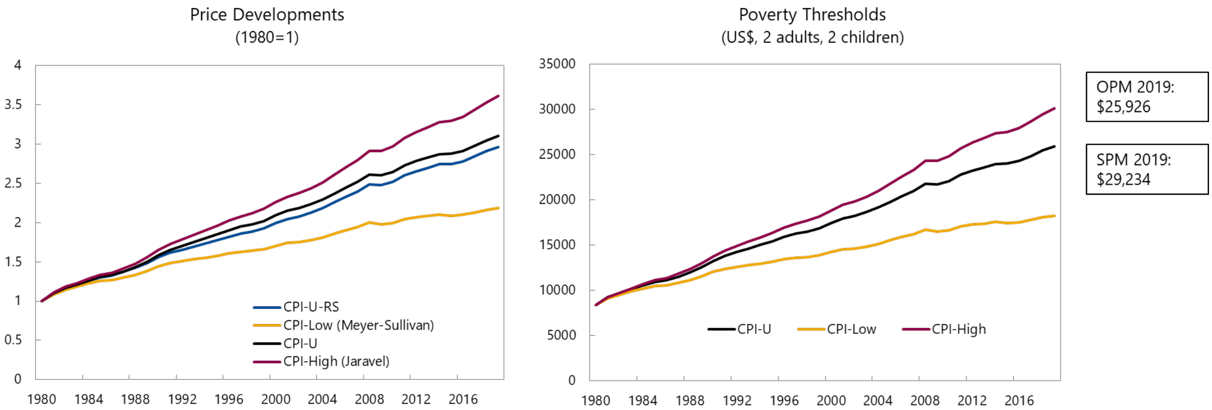
We follow the methodology of Meyer and Sullivan (2012a,b, 2019) in constructing a consumption-based poverty measure and then perform sensitivity analyses to the choice of price index. To convert reported expenditure into a measure of consumption, we make a number of adjustments, which are explained in more detail in Meyer and Sullivan (2012a,b). First, vehicle spending is converted to a service flow equivalent to reflect the value that a consumer receives from owning a car, which is function of a depreciation rate and the vehicle's current market value. Second, we need to impute a rental value of housing expenditure for homeowners and for families living in government or subsidized housing. These calculations involve looking at detailed housing characteristics available in the survey, including for example the number of rooms. Third, we exclude some items that are better interpreted as investment such as healthcare and education and outlays for retirement.[1]

The final step to constructing a consumption-based poverty measure is to specify a poverty threshold. Meyer and Sullivan (2012a,b) calibrate their consumption-based threshold such that 13 percent are poor in 1980, meaning that the poverty rate is the same as the official poverty measure in 1980. The CE survey consists of consumer units (CU) with different numbers of children and adults. A CU in the CE includes all those related by blood and marriage as well as cohabitors who share responsibility for housing, food and other living expenses. To standardize consumption of each unit, an equivalence scale of the form $(A+PK)^F$ is used where A are the number of the adults and K are children, and P and F are set to 0.7. This scale allows for differences in cost between adults and children and exhibits diminishing marginal costs with each additional adult equivalent. The 1980 threshold is then adjusted using different price indexes to obtain thresholds for other years.

2.3 Findings

Figure 1 shows price developments and associated poverty thresholds between 1980 and 2019. Poverty thresholds for a family of four (two children and two adults) in 2019 range from about US$15000 using Meyer and Sullivan's downward adjustment to CPI-U-RS of 0.8 percent every year, to about US$30000 using the 0.4 upward adjustment to CPI-U suggested by Jaravel (2019), the latter is closely aligned with the SPM threshold.

Figure 1: CPI Assumptions and Poverty Thresholds

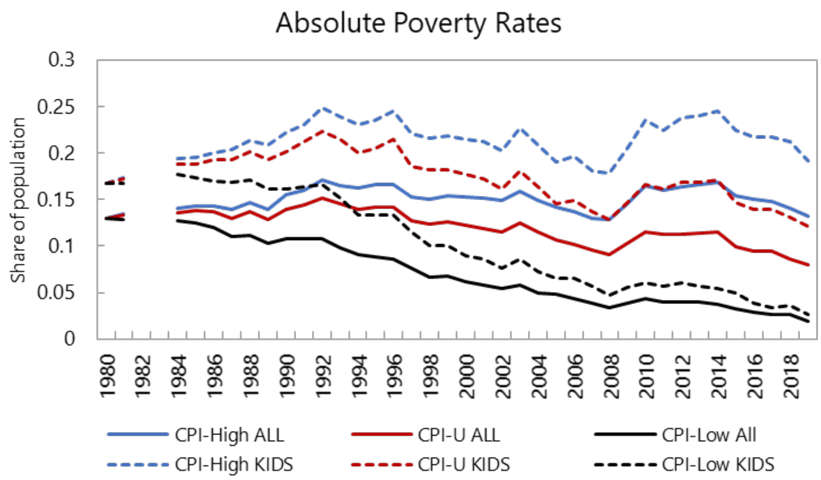


Sources: Meyer and Sullivan (2012a,b, 2013), BLS, CE Survey, CPS, IMF Staff calculations.
Notes: CPI-High= CPI-U+0.4; CPI-Low=CPI-RS-0.8. CPI-High is based on findings by Jaravel (2019), CPI-Low corresponds to the measure used by Meyer and Sullivan (2012b); and Meyer and Sullivan (2019).

[1] We performed robustness checks to check the sensitivity of the results to these assumptions, but found them to only marginally affect the poverty rates. Results are available from the authors upon request.

Using the upwards adjusted CPI-U, there is no decreasing trend in poverty, 2019 rates are close to those in 1980, although there have been up and downward movements (Figure 2). Notably, the Global Financial Crisis led to an upward jump of poverty rates. This differs significantly from Meyer and Sullivan's measure that shows a strong downward trend in poverty rates over the sample period. Looking at child poverty (individuals below 18 years), we find the child poverty rate to be generally above the overall rate, but this difference is much more pronounced for the CPI measure that considers inflation inequality. The reason for this is that non-poor families with children are - on average - significantly closer to the poverty threshold than families without kids. This results in more families with kids (compared to consumer units without kids) becoming poor once the threshold is elevated. According to our measure which does consider inflation inequality, about 20 percent of children lived in poverty in 2019.

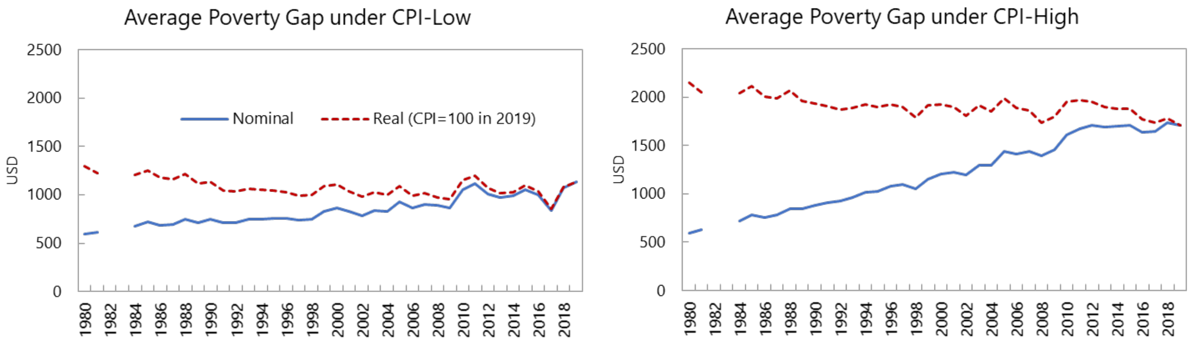
Figure 2: Absolute Consumption Poverty Over The Last Four Decades



Sources: BLS, CE survey, and IMF staff calculations.
Notes: The figure depicts absolute consumption poverty (proportion of population below threshold) using different inflation adjustments. "Kids" are individuals below 18 years. CPI-High= CPI-U+0.4; CPI-Low=CPI-RS-0.8. CPI-High is based on findings by Jaravel (2019), CPI-Low corresponds to the measure used by Meyer and Sullivan (2012b); and Meyer and Sullivan (2019).

We can also examine how the distance to the poverty threshold has evolved, that is compute poverty gaps. The poverty gap is often used by policymakers to inform them how much money is needed to lift the average person (or all poor) out of poverty. Figure 3 shows the average distance of a poor person to the threshold with the downward adjusted CPI-U-RS and the upward adjusted CPI-U to account for inflation inequality, in both nominal and real terms. The 2019 average gap per person was US$1700 for our upward adjusted CPI, and just above US$1000 for Meyer and Sullivan's measure. In nominal terms this is about 2-3 times as high as in the 1980s, but in real terms, the gap has decreased somewhat over time, meaning that on average, poverty was slightly more severe in the 1980s than today.

Figure 3: Absolute Consumption Poverty Gaps



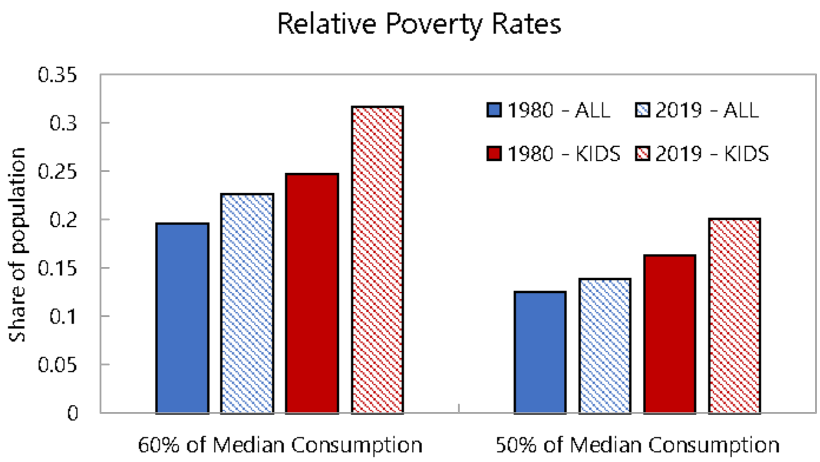
Sources: BLS, CE survey, and IMF staff calculations.
Notes: The figure depicts the average distance of a poor person to the poverty threshold. CPI-High= CPI-U+0.4; CPI-Low=CPI-RS-0.8. CPI-High is based on findings by Jaravel (2019), CPI-Low corresponds to the measure used by Meyer and Sullivan (2012b); and Meyer and Sullivan (2019).

Our consumption based measure that incorporates inflation inequality correlates well with other measures of poverty and deprivation. First, we look at relative poverty rates. We consider two measures, setting the threshold at 50 percent of median consumption (following the OECD methodology) and 60 percent of median consumption (following the EU methodology). Both measures point to higher relative poverty in 2019 compared to 1980, with the increase more pronounced among children, suggesting that an increasing share of Americans are falling behind in terms of how much they can consume (Figure 4). These findings are consistent with studies that have shown declining upward mobility. For instance, Chetty et al. (2017) show that absolute income mobility—the fraction of children who earn more than their parents—has fallen from approximately 90 percent for children born in 1940 to 50 percent for children born in the 1980s.

Second, we plot our measure against income-based measures and other measures of deprivation, such as food insecurity. Both the poverty rate and evolution over time track those other measures more closely than Meyer and Sullivan's downward inflation adjustment (Figure 5). As shown by Shaefer and Rivera (2018), the Survey of Income and Program Participation (SIPP), also shows that between 2003 and 2010, the percentage of households reporting that they fell behind on their mortgage or rent jumped by more than 40 percent, while the percentage of families reporting difficulty paying essential household expenses and meeting medical needs increased by almost 25 percent. While this is broadly in line with our poverty measure that incorporates inflation inequality, it is a different trajectory from Meyer and Sullivan's poverty measure.

As a further robustness test and to explore variation in poverty across states, we adjust for differences in price increases across regions. As shown in Figure 6, the evolution of CPI-U differs with prices rising faster in the northeast and west than in the south and mid-west. Using CPI-U adjusted upwards by 0.4 leads to a range of 2019 poverty thresholds for a family of four (two children and two adults) of about $28,000 in the south and $31,500 in the northeast. To date, there is limited empirical evidence on the differences in consumption across U.S. regions and cities. A notable exception is Diamond and Moretti (2021) who show that low-income residents in the most affordable commuting zone enjoy a level of consumption that is 74 percent higher than that of low-income residents in the most expensive commuting zone. While there are noticeable differences in consumption for high school graduates and high school drop-outs in high and low-cost commuting zones, there is no significant relationship between consumption and cost of living for college graduates.

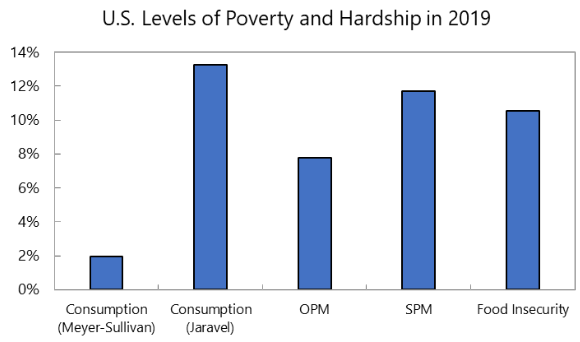
Figure 4: Relative Consumption Poverty

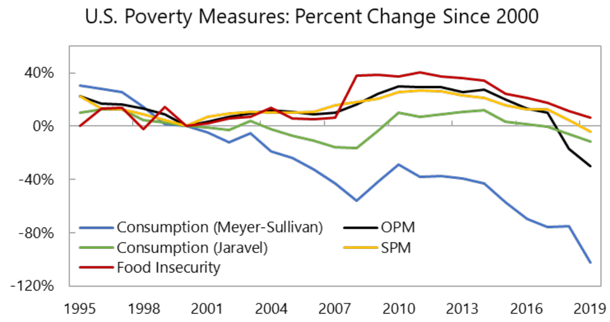


Sources: BLS, CE survey, and IMF staff calculations.
Notes: The chart depicts the share of population with consumption below 60 or 50 percent of median consumption.

Incorporating regional CPIs into the analysis does not significantly affect the overall results. We focus on 1997-2014 as for these years the data coverage is best (with few missing observations on a consumer unit's state of residence). The results show that our overall poverty rates match our bottom up measure using regional poverty rates very closely (Figure 7). But some important regional differences emerge. Poverty in the south was above rates seen in other regions before the GFC, but has recently come down. Following the GFC, the south was no longer the region with the highest poverty rate. The poverty rate in the mid-west increased significantly after the GFC and in most of the post-crisis years was above that in the south. The mid-west is also the only region where poverty did not noticeably decrease in the early 2000s. This is despite the lower consumption thresholds in the mid-west compared to, say, the north-east. Overall, there

Figure 5: Comparison Between Poverty Measures





Sources: BLS, CE survey, Census.gov, USDA, and IMF staff calculations.
Notes: Left side show share of population experiencing absolute consumption or income poverty or food insecurity. Right side shows percent changes in that share over time, across the different measures. Consumption (Meyer-Sullivan) corresponds to CPI-Low, Consumption (Jaravel) to CPI-High.

has been some convergence in poverty rates between 1997-2014 across regions.

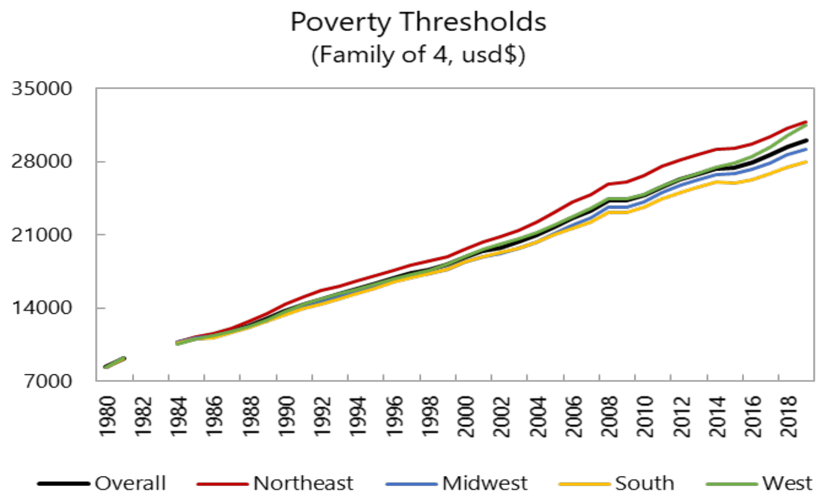
Our findings on poverty rates in the mid-west are in line with a large strand of literature on labor force participation and non-employment. Manufacturing and routine occupations play an important role in the mid-west, but have been increasingly affected by technological changes and offshoring. Austin et al. (2018) find that the economic convergence of American regions has greatly slowed, and rates of long-term non-employment have even been diverging, with the rate of non-employment increasing disproportionately more in the eastern parts of the American heartland.[2] Abraham and Kearney (2020) conclude that labor demand factors, in particular increased import competition from China and the penetration of robots into the labor market, are the most important drivers of observed within-group declines in employment. The fact that the GFC in particular led to a large increase in poverty is in line with the previous literature showing that the vast majority of employment in routine occupations was lost during recessions and never recovered (Jaimovich and Siu, 2020).

2.4 Consumption Poverty And The Expanded Child Tax Credit

The 2021 American Rescue Plan (ARP) significantly expanded the child tax credit (CTC) for 2021. Prior to the ARP, the child tax credit was a $2,000 maximum per child annual credit for families. The exact amount depended on the family's annual income. Individuals with income below $200,000 ($400,000 for joint filers) received a credit of up to $2,000 for every dependent

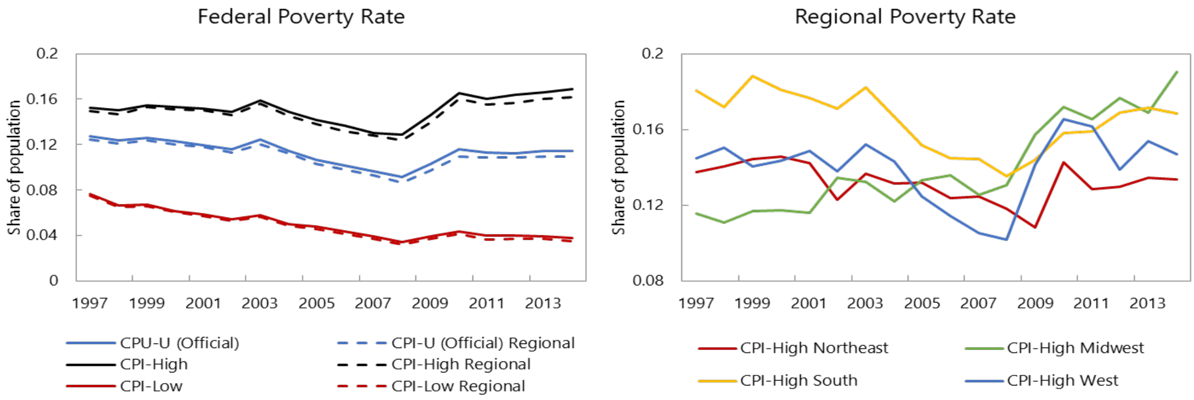
[2] Krueger (2017) shows that labor force participation has fallen more in areas where relatively more opioid pain medication is prescribed, causing the problem of depressed labor force participation and the opioid crisis to become intertwined.

Figure 6: Regional Poverty Thresholds



Sources: BLS, CE survey, and IMF staff calculations.
Notes: Based on CPI-High and family of four.

Figure 7: Regional and Federal Poverty Rates



Sources: BLS, CE survey, and IMF staff calculations.
Notes: CPI-High= CPI-U+0.4; CPI-Low=CPI-RS-0.8. CPI-High is based on findings by Jaravel (2019), CPI-Low corresponds to the measure used by Meyer and Sullivan (2012b); and Meyer and Sullivan (2019).

from ages 0 to 16, and up to $500 for dependants 17 and older. If a taxpayer's income tax liability was less than the maximum value of the child tax credit, the taxpayer may have been eligible to receive all or part of the difference as a refundable credit. The CTC gradually increased, or phased in, as earned income rose above $2,500, with up to $1,400 of the credit refundable per qualifying child for filers with at least $2,500 of earned income. So low income families did not receive the full credit. The ARP made several adjustments: for single filers with income of up to

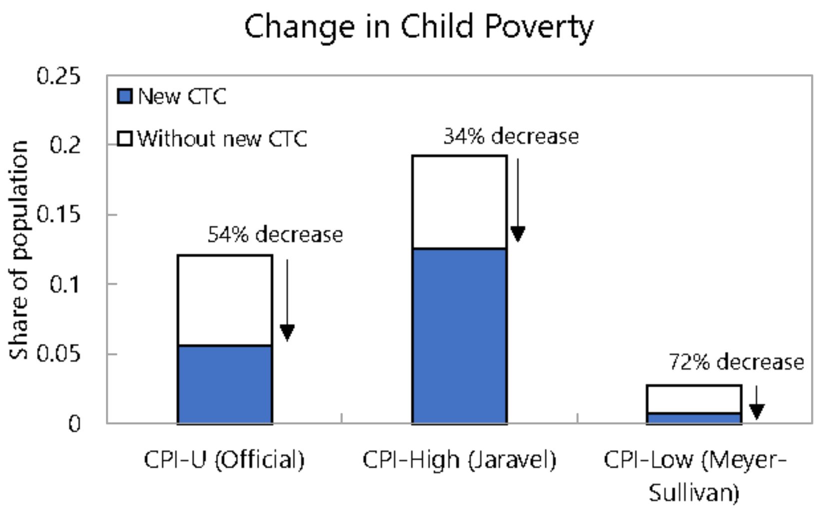
$75,000 ($150,000 for joint filers), the maximum credit increased to $3,600 for dependants from ages 0 to 5 and $3,000 for dependants ages 6 to 17. The increased benefit amount phases out up to an income of $95,000 ($170,000 for joint filers). Moreover, in 2021, the credit amount will be fully refundable for all eligible filers regardless of their earned income amounts. Between July and December 2021, the CTC will be paid out monthly.

A natural question is whether the CTC expansion can significantly lower child consumption poverty. We have consumer unit annual income, marital status and information on the number and age of children in the CE survey and use that with the CTC income thresholds to calculate the additional CTC that families get under the ARP compared to the old system. In order to calculate the effect on consumption poverty, we simply add the additional CTC to the consumption of the consumer unit and test whether the consumer unit - and therefore its kids - is still below the poverty threshold. This exercise is based on two assumptions. First, we assume that the whole increase of the CTC translates into higher consumption. Since this exercise focuses on families close to the poverty threshold, the propensity to spend should be significantly higher than for the rest of the population. That said, an important caveat to this static exercise is that it abstracts from behavioral changes in response to the increase in the CTC, including for instance, employment decisions of the household head. Thus, our estimates provide an upper bound to the possible impact. Second, we assume that all married couples are filing jointly.

The CTC expansion leads to a significant decrease in poverty among households with children, with some variation depending on which CPI is used. Using the official CPI, the CTC is projected to reduce child poverty from around 12 percent to 5.5 percent, which is a 54 percent reduction (Figure 8). Using the CPI that incorporates income inequality, the reduction is around 34 percent. These results are broadly similar in magnitude to recent estimates from Corinth et al. (2021) and Columbia University's Center on Poverty and Social Policy, using income-based poverty.[3]

[3] https://www.povertycenter.columbia.edu/news-internal/2021/presidential-policy/biden-economic-relief-proposal-poverty-impact

Figure 8: The Child Tax Credit And Consumption Poverty



Sources: BLS, CE survey, and IMF staff calculations.
Notes: CPI-High= CPI-U+0.4; CPI-Low=CPI-RS-0.8. CPI-High is based on findings by Jaravel (2019), CPI-Low corresponds to the measure used by Meyer and Sullivan (2012b); and Meyer and Sullivan (2019).

3 The Determinants of Poverty

3.1 Data and Methodology

3.1.1 Data

For the following analyses, we use the overall poverty rate adjusted for inflation inequality and differences in regional inflation from 1984 to 2019 (CPI-High Regional in Figure 7) as the dependent variable. In total, we have 870,047 observations (98,997 observations for the poor and 771,050 for the non-poor), with an average of 24,168 observations per year and a minimum of 17,149 observations in year 1985 and maximum of 32,326 observations in year 2003. There are missing observations for some household characteristics (such as "receiving housing support", "hours worked per week by CU head", "occupation of CU head"). Therefore, to carry out robustness checks, we use subsets of the original data (783,403 observations for "housing support", 632,231 for "hours worked", 598,119 for "occupation"). We provide a brief discussion on the results associated with these robustness checks in Appendix B.

A closer look at the data reveals the main household characteristics associated with our poverty measure. As shown in Table 1, poor consumption units tend to be larger with more children and headed by younger individuals. Family size among poor consumption units has increased over the years in contrast to the non-poor. In addition, poor consumption units are more likely to be single headed rather than headed by a married couple. Single parents are mostly single mothers.

Poor consumption units are more likely to be headed by less educated individuals that are not in full-employment and are more likely to be minorities. Additionally, a larger fraction of the poor receive housing support compared to the non-poor. The data also suggest a somewhat higher prevalence of poverty in the south and mid-west compared to the northeast and west even after adjusting for price differences.

Table 1: Who are the poor?

	1984-1989		1990-1999		2000-2007		2008-2019	
Variable	Poor	Not poor	Poor	Not poor	Poor	Not poor	Poor	Not poor
Family size	2.95	2.54	3.10	2.48	3.16	2.48	3.14	2.39
No. children <18	1.17	0.64	1.25	0.64	1.19	0.62	1.12	0.54
No. adults >64	0.29	0.29	0.26	0.31	0.27	0.32	0.29	0.38
Age of CU head	42.40	46.68	42.54	48.05	43.33	49.37	44.75	51.58
Single parent	16.8%	5.0%	15.5%	5.3%	12.1%	5.2%	10.6%	4.7%
Single mother	16.1%	4.3%	14.9%	4.6%	11.2%	4.4%	9.6%	3.9%
Married	35.5%	58.2%	38.7%	56.4%	40.5%	55.4%	38.9%	52.6%
Widowed	14.0%	10.9%	11.2%	10.8%	9.8%	10.2%	8.3%	9.7%
Separated/divorced	18.3%	14.4%	19.1%	16.2%	17.3%	16.6%	18.0%	17.8%
Never married	32.1%	16.5%	30.9%	16.7%	32.4%	17.8%	34.8%	19.9%
High school or less	53.6%	20.2%	42.9%	15.2%	38.1%	12.3%	32.6%	9.8%
Some college	42.0%	54.3%	52.1%	56.6%	56.0%	57.5%	58.8%	54.2%
Bachelor's or more	4.3%	25.5%	5.0%	28.2%	5.9%	30.2%	8.5%	36.0%
White	67.0%	87.3%	69.0%	85.3%	71.3%	82.9%	71.8%	80.3%
Black	28.4%	8.9%	25.1%	10.0%	21.5%	10.1%	19.3%	10.9%
Hispanic	14.1%	4.7%	18.2%	6.0%	24.0%	8.5%	25.8%	11.6%
Northeast	22.0%	22.2%	22.2%	20.2%	17.2%	17.3%	16.8%	18.8%
Midwest	26.6%	25.5%	22.8%	25.3%	24.0%	22.8%	24.5%	21.3%
South	34.9%	27.8%	35.0%	30.6%	38.3%	33.3%	37.8%	35.9%
West	16.4%	24.4%	20.0%	23.9%	23.0%	25.5%	20.9%	24.1%
Receives housing support[1]	13.3%	2.5%	13.6%	3.3%	16.8%	4.9%	16.1%	6.8%
Full time worker[1]	67.9%	88.9%	72.9%	87.3%	69.9%	83.7%	65.1%	82.2%
No. observations	12,215	100,583	25,161	183,083	27,845	216,120	33,776	271,264

Notes: This table shows average values for the respective groups (poor vs. not poor) and time periods.
[1] Statistics based on a restricted sample of 783,403 (Receives housing support) and 632,231 (Full time worker) observations, respectively.

3.1.2 Motivation Behind Empirical Methods

This paper provides a comprehensive analysis on the drivers of poverty via the application of various estimation methods, such as probit, tobit, and machine learning algorithms. While it is widely recognised in the poverty literature that traditional methods such as probit or tobit can contribute to understanding the main factors influencing poverty status, the application of machine learning techniques in this context is limited to only a few countries (Sohnesen and Stender, 2016; Zhao et al., 2019). Because machine learning, unlike probit or tobit, offers a nonparametric and flexible way to model poverty status, we hope to develop a superior poverty-identification technique that not only results in higher predictive power but also reveals more clearly the underlying empirical relationships considered.

In the following sections, we describe in more detail the application of the aforementioned approaches.

3.1.3 Probit

In this section, we present the econometric model used to identify the determinants of poverty status. Based on our preferred poverty measure, we represent poverty status as a binary variable Y which takes the value of 1 if a given consumer unit is poor and 0 otherwise. It is widely understood that a linear model would not be adequate in this scenario as it assumes the conditional probability function to be linear, thus we adopt a probit model which is commonly used in the poverty literature (Bahta and Haile, 2013; Cho and Kim, 2017; Drescher and Janzen, 2021).
Our baseline model is

$$Y = \alpha_1 + \boldsymbol{X}\boldsymbol{\beta_1} + \boldsymbol{H}\boldsymbol{\gamma_1} + \boldsymbol{T}\boldsymbol{\delta_1} + \epsilon_1. \tag{1}$$

In probit regressions, the cumulative normal distribution function $\Phi(\cdot)$ is used to model the regression function when the dependent variable is binary (Stock and Watson, 2015). Therefore we have

$$\mathbb{E}(Y|\boldsymbol{X}, \boldsymbol{H}, \boldsymbol{T}) = \mathbb{P}(Y = 1|\boldsymbol{X}, \boldsymbol{H}, \boldsymbol{T}) = \Phi(\alpha_1 + \boldsymbol{X}\boldsymbol{\beta_1} + \boldsymbol{H}\boldsymbol{\gamma_1} + \boldsymbol{T}\boldsymbol{\delta_1}), \tag{2}$$

where the probability of being poor is modelled as a function of CU characteristics $\boldsymbol{X}$ (number of children, number of elderly, whether the CU is headed by a single mother, race, ethnicity, region), characteristics of the CU head $\boldsymbol{H}$ (education, marital status), $t-1$ indicator variables in $\boldsymbol{T}$ that control for the survey year ($t = 36$), and ϵ_1 is an error term assumed to be distributed $\sim \mathcal{N}(0, 1)$. Table C.1 provides a detailed description of the variables considered.

Since the dependent variable Y is a nonlinear function of all covariates in $\boldsymbol{X}$,$\boldsymbol{H}$, and $\boldsymbol{T}$, the estimated coefficients in $\boldsymbol{\beta_1}$,$\boldsymbol{\gamma_1}$, and $\boldsymbol{\delta_1}$ have no simple interpretation. In order to quantify the effect of a given variable on the probability of being poor, we calculate average marginal effects (AME) with respect to each characteristic we are interested in (discussed in Section 3.2.1). To evaluate the time dynamics of $\boldsymbol{\beta_1}$ and $\boldsymbol{\gamma_1}$, we interact all covariates in $\boldsymbol{X}$ and $\boldsymbol{H}$ with a categorical variable that takes values 1,2,3,4 according to different time periods (1984-1989, 1990-1999, 2000-2007, 2008-2019, respectively).

3.1.4 Tobit

One limitation with respect to the probit model defined in Equation 2 is that we are forgoing valuable information regarding the poor, namely how deep into poverty they are. For this reason, we adobt a tobit model in Equation 3, coined by Tobin (1958), and following Asogwa et al. (2012), in order to incorporate the severity of poverty and to estimate the determinants of poverty depth. First, we define our latent dependent variable P^* as $[p_i^*] = \left(\frac{z_i - c_i}{z_i}\right)$, where z_i and c_i represent the poverty threshold and consumption for CU i in USD, respectively. This measure of poverty depth represents how far away a CU is from the poverty line, in proportion to the poverty threshold itself. Therefore,

$$P = \begin{cases} \alpha_2 + \boldsymbol{X\beta_2} + \boldsymbol{H\gamma_2} + \boldsymbol{T\delta_2} + \epsilon_2 \equiv P^*, & \text{if } P^* \in (0,1) \\ 0, & \text{if } P^* \leq 0, \end{cases} \tag{3}$$

where P is discrete and equal to 0 when the CU is not poor and continuous within the range (0,1) when the CU is poor, and ϵ_2 is an independently distributed error term assumed to be normal with mean zero and constant variance σ^2.

Although the estimated coefficients from Equation 3 can only be interpreted as the respective marginal effects on poverty depth for all observations (poor and non-poor CUs), we follow the decomposition method by McDonald and Moffitt (1980) to gain additional insight on how these characteristics affect poverty depth among the poor. As shown in Equation 4, the total change in P with respect to any variable, let's take X_1, can be disaggregated into two components: (1) the intensive margin, i.e. the change in the expected poverty depth with respect to X_1 conditional on being poor, weighted by the probability of being poor; and (2) the extensive margin, i.e. the change in the probability of being poor with respect to X_1, weighted by the expected poverty depth among the poor. This decomposition illustrates the advantages the tobit model offers over the probit model, as the latter only focuses on the extensive margin.

$$\frac{\partial \mathbb{E}[P]}{\partial X_1} = \mathbb{P}(0 < P < 1)\frac{\partial \mathbb{E}[P|0 < P < 1]}{\partial X_1} + \mathbb{E}[P|0 < P < 1]\frac{\partial \mathbb{P}(0 < P < 1)}{\partial X_1} \tag{4}$$

Similarly to our probit analysis, we evaluate the time dynamics of $\boldsymbol{\beta_2}$ and $\boldsymbol{\gamma_2}$ by interacting all covariates in $\boldsymbol{X}$ and $\boldsymbol{H}$ with a categorical variable that takes values 1,2,3,4 according to different time periods (1984-1989, 1990-1999, 2000-2007, 2008-2019, respectively).

3.1.5 Machine Learning

We begin this exercise by considering a standard random forest model (Breiman, 2001). This supervised algorithm can be used for classification, i.e. to determine whether a consumer unit is poor based on a set of observables. In our case, given data on poverty status and household characteristics in $\boldsymbol{X}$ and $\boldsymbol{H}$, the random forest consists of various classification trees that attempt to predict whether a given household is poor or not (see Figure A.1 for an example). Finally, these decisions are merged together to get a more accurate and stable prediction. One limitation associated with traditional random forest is that it does not perform very well with respect to unbalanced panels. In our case, roughly 11.4% of the consumer units in our sample are poor, leading us to explore variations of traditional random forest (RF) models, such as balanced random forest (BRF)[4] and gradient boosting (GBM)[5]. While the performance of these models with respect to traditional random forest varies depending on specific circumstances, researchers have found convincing evidence in favour of BRF models, when the dataset is heavily tilted towards one group (Ling and Li, 1998; Drummond and Holte, 2003). More detail on the application of these models in the context of our analysis is provided in Appendix A.

3.1.6 Assessing Model Performance

To assess how well we can predict poverty status, we train our probit and machine learning models on a training sample (roughly 70% of the original data)[6] and compare the out-of-sample classification error. This makes the performances of the individual models comparable as other metrics, such as R-squared, do not exist in the same way across our approaches. In general, we focus on Type II errors, i.e. how many people we fail to classify as poor.

3.2 Findings

3.2.1 Probit

Table 2 presents the estimated coefficients in $\boldsymbol{X}$ and $\boldsymbol{H}$ resulting from our probit analysis. As previously mentioned, the magnitude of these coefficients is not easy to interpret but the sign is indicative of whether a given variable is positively or negatively correlated with the likelihood of being poor. In order to correctly interpret the coefficients on variables related to region (West, South, Midwest), marital status (Married, Widowed, Separated or divorced) and education (Some college, Bachelor's or higher), one must relate them to the respective omitted category (Northeast, Never married, Less than high school). For example, a negative sign on

[4] This method, as the name suggests, requires that each tree is given a sample where the proportion of poor vs. not poor is 1:1.

[5] Boosting is a general technique that improves accuracy by training subsequent trees (or learners) on the errors made by previous trees.

[6] Training (test) sample is constructed by randomly choosing 70% (30%) of households each year.

the coefficient for "West" means that, holding all other factors constant, a consumer unit is less likely to be poor if it is located in the West rather than Northeast. It is worthwhile to note that almost all coefficients are statistically significant at the 1% level. In addition to the two specifications discussed in Section 3.1.3, we provide a robustness check (3) where we control for the age of the CU head using a second degree polynomial. We observe a modest shrinkage in the coefficients related to marital status, which can be explained by the positive correlation of age and being married, widowed, separated or divorced. After controlling for age of the CU head as well as marital status, the coefficient on "Single mother" becomes insignificant.

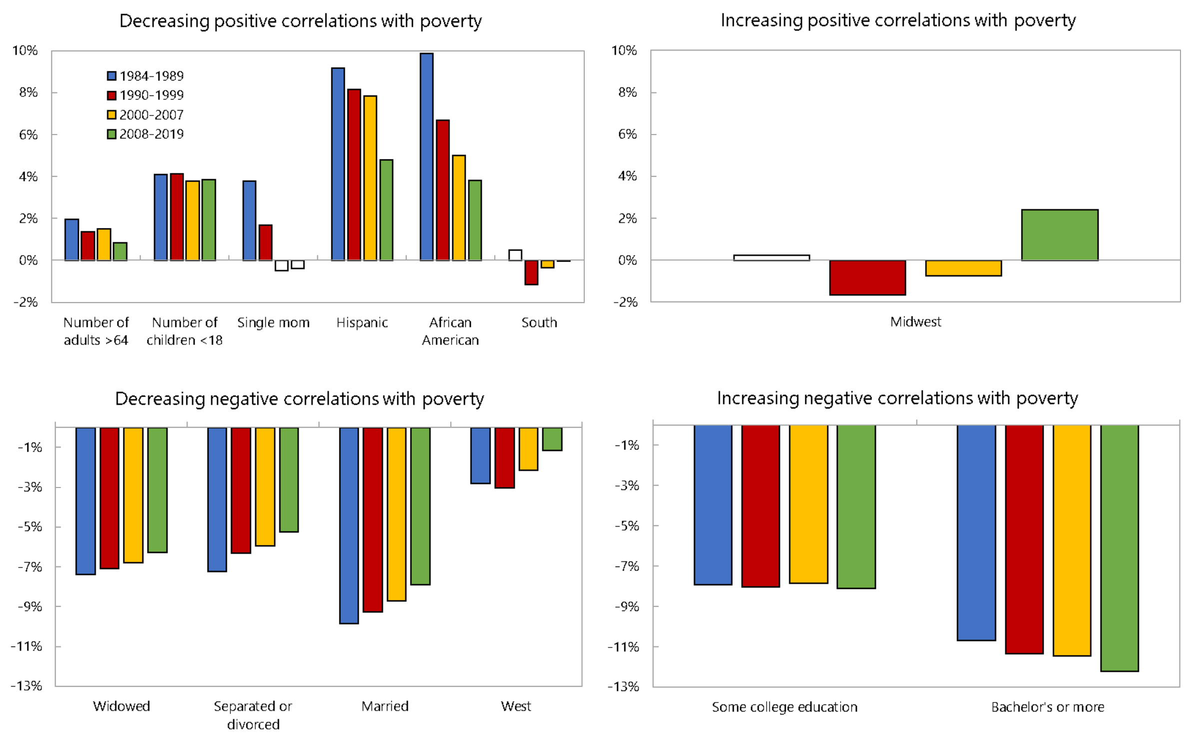
Table 2: Probit - Estimated coefficients

	(1)				(2)	(3)
Variable	1984-1989	1990-1999	2000-2007	2008-2019	Pooled sample	Pooled sample
No. adults >64	0.123***	0.085***	0.093***	0.053***	0.080***	0.083***
No. children <18	0.257***	0.258***	0.236***	0.241***	0.244***	0.226***
Single mother	0.213***	0.100***	−0.032*	−0.025	0.042***	0.027
Some college	−0.667***	−0.619***	−0.583***	−0.584***	−0.607***	−0.637***
Bachelor's or higher	−1.304***	−1.326***	−1.262***	−1.257***	−1.279***	−1.264***
Widowed	−0.648***	−0.599***	−0.563***	−0.503***	−0.549***	−0.351***
Separated or divorced	−0.625***	−0.504***	−0.462***	−0.390***	−0.455***	−0.162***
Married	−0.948***	−0.777***	−0.696***	−0.599***	−0.704***	−0.451***
Hispanic	0.461***	0.419***	0.407***	0.266***	0.357***	0.353***
African American	0.491***	0.355***	0.276***	0.216***	0.303***	0.366***
West	−0.193***	−0.210***	−0.145***	−0.073***	−0.140***	−0.163***
Midwest	0.015	−0.110***	−0.047***	0.141***	0.010	−0.011
South	0.029	−0.074***	−0.022*	−0.002	−0.024***	−0.044***
Control for age of CU head	No				No	Yes
Pseudo R squared	0.171				0.168	0.189
No. observations	609,303				609,303	609,303

Notes: This table shows the estimated coefficients in $\boldsymbol{\beta_1}$ and $\boldsymbol{\gamma_1}$ from Equation 2, as well as time dynamics. Additionally, we provide a robustness check by controlling for the age of the reference person by including second degree polynomial. ***, **, and * represent 1%, 5%, and 10% levels of significance, respectively.

In order to quantitatively assess the effect of each of the variables considered on the probability of being poor, i.e. the extensive margin, we calculate the respective average marginal effects (AMEs), to not confuse with marginal effects at the mean. In other words, for each variable we evaluate the marginal effect for each consumer unit in our sample, and simply take the unweighted average. Figure 9 shows the time dynamics of the AMEs associated with specification (1) in Table 2. We observe that African American and Hispanic households, as well as households with more children, are more likely to be poor. However, notably, the positive correlation of race and ethnicity with respect to poverty status has halved since 1980's. More specifically, we find that, after controlling for other factors, Hispanic and African American families in the 80's were 9-10% more likely to be poor than their counterparts (non-Hispanic, non-African American CUs), while now they are only 4-5% more likely to be poor. This could reflect successful policies aimed at reducing racial and ethnic inequality but also stresses the point that there is still more work to do. On the other hand, consumer units where the household head has at least a Bachelor's degree are roughly 12% less likely to be poor than households where the reference person has not completed high school. Similarly, after controlling for other household characteristics, we find that CUs where both adults are married are about 8% less likely to be poor than households headed by an individual who has never been married. While the importance of education in determining poverty status has increased over time, we find the opposite with respect to marital status.

Figure 9: Probit - Estimated Marginal Effects



Sources: BLS, CE survey, and IMF staff calculations.
Notes: This chart shows time dynamics of average marginal effects with respect to the probability of being poor, for each of the variables considered in our baseline specification. All effects are statistically significant at the 1% level, except for shaded columns (not significant).

We also find that, even after controlling for marital status, education, and number of children, single mother households do not necessarily experience a higher likelihood of being poor in the last twenty years. Though the opposite can be said regarding the first fifteen years of our sample, where single mother households are positively, and significantly, correlated with poverty. While the role of education has become more important over time, our results suggest that marital status played a bigger role in the 1980s than in present time. The result that education has become a more important predictor of poverty is in line with a number of papers by Case and Deaton (e.g., Case and Deaton (2020)), which find a central division is between those who do or do not have a four-year college degree in explaining wage and health outcomes. Finally, controlling for inflation inequality across regions and all other factors, we find that following the Global Financial Crisis poverty became more prevalent in the Midwest, compared to the rest of the United States.

3.2.2 Tobit

Table 3 presents the decomposition of the estimated coefficients from Equation 3. With the exception of "Midwest" and "Single mother", all of the coefficients are statistically significant at the 1% level. Focusing on the overall margins, a positive sign would translate into an increase in poverty depth, while a negative sign would represent a decrease in the distance to the poverty threshold. A general comparison with our probit results (Table 2, (2)) shows that the variables considered have similar economic relationships with both poverty depth and poverty status, as expected.

The very intuitive formula outlined in Equation 4 allows us to comment on multiple findings. The fraction of the overall margin that is due to the intensive margin is roughly equal to 0.17. This can be calculated by taking the ratio of the intensive margin to the overall margin in Table 3. This means that for each of the factors considered, there is a strong effect on the overall probability of being poor, and a much weaker effect on poverty depth among the poor. Confirming our probit findings, minority and larger CUs, especially with more children, are more likely to be poor while consumer units where the adults are married and the head has achieved higher education are least likely to be poor.

It is helpful to think about the magnitude of intensive margin in relation to the distribution of poverty depth, shown in Figure 10. With a median of roughly 0.20, a representative, poor consumer unit is typically 20% into poverty, though there is a high concentration of CUs just below the threshold. While region does not play a big role, households where the adults are married and at least one parent has achieved a Bachelor's degree are more likely to be closer to the poverty line. For those poor consumer units that are on the margin, i.e. very close to the poverty line, a decision such as pursuing a college degree or getting married, holding everything

Table 3: Tobit - Decomposition of Estimated Overall Margins (Pooled Sample)

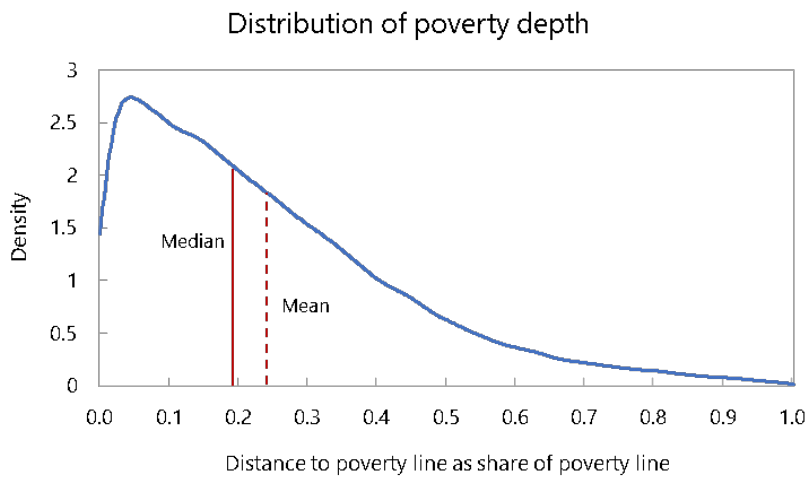
	Overall	Intensive	Extensive
No. adults >64	0.023***	0.004***	0.019***
No. children <18	0.090***	0.014***	0.076***
Single mother	−0.006*	−0.001*	−0.005*
Some college	−0.235***	−0.038***	−0.197***
Bachelor's or higher	−0.523***	−0.084***	−0.439***
Widowed	−0.244***	−0.039***	−0.205***
Separated or divorced	−0.204***	−0.033***	−0.171***
Married	−0.308***	−0.049***	−0.259***
Hispanic	0.134***	0.021***	0.113***
African American	0.111***	0.018***	0.093***
West	−0.061***	−0.010***	−0.051***
Midwest	0.000	0.000	0.000
South	−0.014***	−0.002***	−0.012***
Pseudo R squared		0.193	
No. observations[1]		609,214	

Notes: This table shows the estimated overall margins in $\boldsymbol{\beta_2}$ and $\boldsymbol{\gamma_2}$ from Equation 3 and the decomposition into intensive and extensive margins.

[1] 89 observations were dropped because consumption was either zero or negative.

else constant, could have significant positive implications. In more practical terms, we find that, after controlling various household characteristics, poor CUs where the reference person has at least a Bachelor's degree, as opposed to not having graduated high school, will, on average, be roughly 8.5% closer to the poverty threshold. Similarly, a poor household where the reference person is married, as opposed to single, will, on average, be 5% closer to the poverty line. On the other hand, poor African American and Hispanic households are more likely to be deeper into poverty, though the magnitude of the intensive margins are 2-3 times smaller than other variables such as marital status and education.

Figure 10: Tobit - Distribution of Dependent Variable Among the Poor



Sources: BLS, CE survey, and IMF staff calculations.
Notes: This chart shows the distribution of the dependent variable when the household is poor. We define poverty depth as the distance to poverty line divided by the the poverty line, hence our dependent variable becomes a ratio. The closer the value is to 0, the closer the CU is to the poverty threshold, and vice-versa.

Table 4 shows time dynamics for the coefficients in $\boldsymbol{X}$ and $\boldsymbol{H}$ discussed above. We can observe similar patterns to Section 3.2.1 where the overall effects on race, ethnicity and marital status shrink over time. Additionally, consumer units located in the Midwest following the Global Financial Crisis are, on average, deeper into poverty than CUs in the rest of the United States. Although the overall margins are getting smaller over time, it is worthwhile to note that the estimated probability of being poor, $\mathbb{P}(0 < P < 1)$, i.e. the fraction of the overall margin associated with the intensive margin, has increased over time by roughly 2% (from 15% in 1984-1989 to 16.9% in 2008-2019). This implies that the role of household and socio-economic characteristics in determining whether a CU is poor (i.e. extensive margin) has decreased marginally over time and, consequently, they now play a bigger role in explaining poverty depth.

Table 4: Tobit - Time Dynamics in Estimated Overall Margins

	1984-1989	1990-1999	2000-2007	2008-2019
Fraction of intensive margin	0.150	0.153	0.161	0.169
No. adults >64	0.035***	0.024***	0.028***	0.013***
No. children <18	0.094***	0.095***	0.088***	0.088***
Single mother	0.048***	0.010	−0.036***	−0.025***
Some college	−0.267***	−0.237***	−0.221***	−0.223***
Bachelor's or higher	−0.544***	−0.541***	−0.512***	−0.507***
Widowed	−0.287***	−0.262***	−0.251***	−0.221***
Separated or divorced	−0.272***	−0.222***	−0.208***	−0.175***
Married	−0.410***	−0.335***	−0.305***	−0.260***
Hispanic	0.178***	0.155***	0.155***	0.096***
African American	0.186***	0.132***	0.098***	0.075***
West	−0.087***	−0.090***	−0.060***	−0.032***
Midwest	0.000	−0.053***	−0.023***	0.056***
South	0.011	−0.034***	−0.013***	−0.006
Pseudo R squared		0.197		
No. observations[1]		609,214		

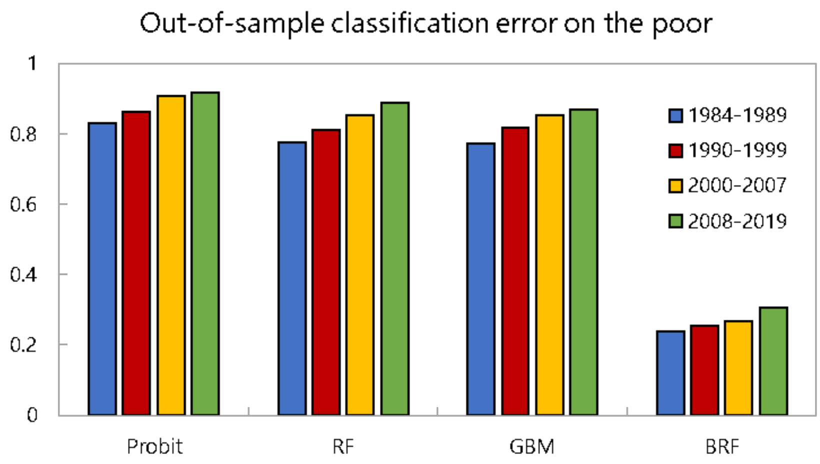
Notes: This table shows estimated time dynamics in $\boldsymbol{\beta_2}$ and $\boldsymbol{\gamma_2}$ from Equation 3 and, for each time period, the fraction of the overall margin due to the intensive margin.

[1] 89 observations were dropped because consumption was either zero or negative.

3.2.3 Machine Learning

We first compare the predictive accuracy of our machine learning models in order to select our preferred method of analyzing the determinants of poverty. Figure 11 compares the out-of-sample classification error on poor consumer units for each of the methods considered in this paper, including probit. We can see that, given the nature of our dataset, standard machine learning does not provide an advantage over more traditional methods that have been extensively relied on in the poverty literature. Though, by down-sampling our data (i.e. forgoing information on the "not poor" such that trees in random forest are trained on a balanced sample), we can achieve significant improvements in prediction power. More specifically, by adopting a balanced random forest algorithm (BRF) we can improve the classification error on the poor by roughly 65%. A rather sizeable classification error of 25% prevails given our preferred method of estimation and cannot be ignored, as there may be additional factors that are not taken into account.

Figure 11: Comparison of Predictive Performance

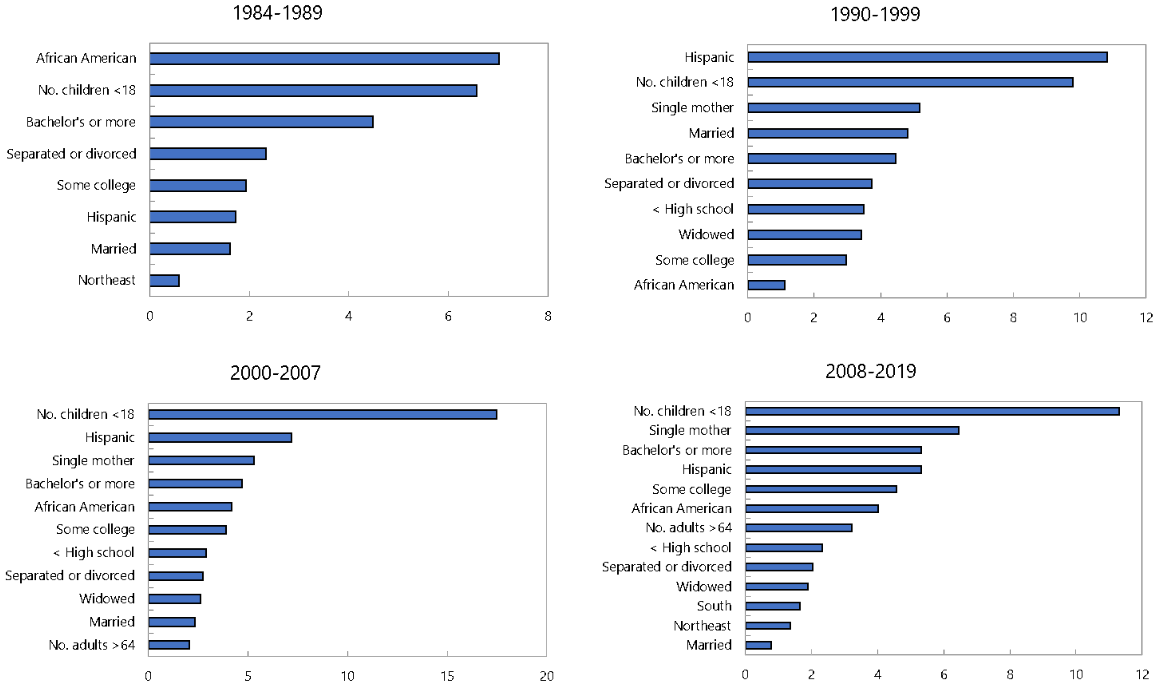


Sources: BLS, CE survey, and IMF staff calculations.
Notes: Out-of-sample prediction was done on a 30% split of the original data.

Figure 12 presents the ranking of variables in $\boldsymbol{X}$ and $\boldsymbol{H}$ in terms of their contribution to the predictive accuracy associated with the balanced random forest algorithm. Although the interpretation of these results compared to probit or tobit differs, we should expect to see similar patters. In other words, if a variable has a substantial effect on the likelihood of being poor, it is likely that it will also be an important factor for prediction.

While regional differences do not seem to contribute much, the number of children and having a Bachelor's degree or more are important predictors of poverty, and have become more so over time. On the other hand, while being an African American household was the best identifier of poverty in the 1980's, its relative importance has decreased substantially. The same applies to Hispanic heritage since the 1990's. Although marital status of the household head has a relatively small contribution to the predictive accuracy of our model, whether a household is headed by a single mother proves to be a major identifier of poverty status. This result, though a bit contradictory of the discussion in Section 3.2.1 and 3.2.2, can be explained by the nature of the random forest algorithm. Generally speaking, for a given tree, at each decision point the objective is to find a variable that leads to the optimal split of the data. Since the fraction of poor CUs that are lead by a single mother is roughly four times greater than for non-poor consumer units, this variable will have a high likelihood of being picked.

Figure 12: Balanced Random Forest - Variable Importance



Sources: BLS, CE survey, and IMF staff calculations.
Notes: This charts shows, for each time period, the ranking of variables based on their contribution to the predictive accuracy of balanced random forest (BRF).

4 Conclusion

A substantial literature has examined the evolution of poverty in the United States. In this paper, we use micro-data from household expenditure surveys to measure consumption poverty over the last four decades. Our results document that the share of the population in poverty is highly sensitive to the choice of price index. In contrast to previous studies based on consumption poverty, we show that when accounting for inflation inequality, there is no clear downward trend in poverty rates. Moreover, we show that child poverty has increased over the last forty years. Our preferred measure of poverty correlates well with other measures of material hardship, such as food insecurity and difficulties with paying rent, meeting medical needs and paying for essential household expenses. Relative consumption poverty measures illustrate that an increasing share of Americans are falling behind.

We also shed new light on the determinants of U.S. poverty, using probit, tobit, and machine learning models. Families with more children, African American families, and Hispanic families are more likely to be poor. On the other hand, families where the household head holds a Bachelor's degree or more or where the adults are married are less likely to be poor. Controlling for time dynamics, regional differences have widened since the Global Financial Crisis (GFC), with families located in the mid-west experiencing a higher likelihood of poverty. Moreover, a Bachelor's degree became even more important in the last decade compared to the 1980s while marital status and race became less important in explaining poverty. However, race and ethnicity are still significant determinants of poverty, even after we control for many other characteristics, such as education and employment status. Machine learning models confirm that we are focusing on the most important explanatory variables. However, even with our preferred model specification a sizeable prediction error remains, pointing to the role of random factors, or "bad luck" in explaining poverty.

While more than 15 percent of adults and 20 percent of children lived in poverty in 2019, the good news is that policies can play an important role in combating economic hardship. A natural experiment with the expanded child tax credit that was put in place for one year in the 2021 American Rescue Plan shows that it can reduce annual child poverty by more than a third. Our findings on the still significant role of race/ethnicity in determining poverty also point to the importance of continuing with policies that advance racial equity and support for under-served communities. Finally, in line with previous studies, we document the important role of education in reducing the likelihood of falling into poverty. Investment in early childhood has clear benefits, while there is increasing evidence that there is a divide between those with and without a college degree in terms of income and health (Case and Deaton, 2020).

There are several avenues for future work. First, our methodologies could be applied to study poverty during the COVID pandemic and assess how effective the fiscal stimulus measures such as the stimulus checks and unemployment benefit programs were in combating poverty. This could yield useful lessons for future crises. Another interesting direction for future research is to explore how the consumption patterns of the poor have evolved and adjust for changes in living standards and the quality of the underlying consumption basket. Items that were not available or needed in the 1980s, may now be considered necessities (such as a computer to work from home). Other countries (e.g., New Zealand) have attempted to identify a minimum consumption basket that changes through time as new items become essential. This could provide an interesting alternative to standard poverty measures. The poor may also have reduced their consumption of public goods, such as education, especially when adjusting for the decline in the quality of first and secondary public education. This would have important implications not only for present levels of poverty but also future poverty rates.

References

Abraham, K. G. and M. S. Kearney (2020): "Explaining the Decline in the US Employment-to-Population Ratio: A Review of the Evidence," *Journal of Economic Literature*, 58, 585–643.

Argente, D. and M. Lee (2021): "Cost of Living Inequality During the Great Recession," *Journal of the European Economic Association*, 19, 913–952.

Asogwa, B. C., J. Umeh, and V. Okwoche (2012): "Estimating the Determinants of Poverty Depth among the Peri–Urban Farmers in Nigeria," *Current Research Journal of Social Sciences*, 4, 201–206.

Austin, B., E. Glaeser, and L. Summers (2018): "Jobs for the Heartland: Place-Based Policies in 21st-Century America," *Brookings Papers on Economic Activity*, 2018, 151–232.

Bahta, Y. T. and B. O. Haile (2013): "Determinants of Poverty of Zoba Maekel of Eritrea: A Household Level Analysis," *International Journal of Food and Agricultural Economics*, 1, 73–84.

Berndt, E. R. (2006): "The Boskin Commission Report After a Decade: After-life or Requiem?" *International Productivity Monitor*, 12, 61–73.

Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

Case, A. and A. Deaton (2020): *Deaths of Despair and the Future of Capitalism*, Princeton University Press.

Chetty, R., D. Grusky, M. Hell, N. Hendren, R. Manduca, and J. Narang (2017): "The Fading American Dream: Trends in Absolute Income Mobility Since 1940," *Science*, 356, 398–406.

Cho, S. and T. Kim (2017): "Determinants of Poverty Status in Rwanda," *African Development Review*, 29, 337–349.

Corinth, K., B. D. Meyer, M. Stadnicki, and D. Wu (2021): "The Anti-Poverty, Targeting, and Labor Supply Effects of the Proposed Child Tax Credit Expansion," *NBER Working Paper*, 29366.

Diamond, R. and E. Moretti (2021): "Where is Standard of Living the Highest? Local Prices and the Geography of Consumption," *NBER Working Paper*, 29533.

Drescher, K. and B. Janzen (2021): "Determinants, persistence, and dynamics of energy poverty: An empirical assessment using German household survey data," *Energy Economics*, 102.

Drummond, C. and R. C. Holte (2003): “C4.5, Class imbalance, and cost sensitivity: why understanding beats oversampling,” *Proceedings of the Twentieth International Conference on Machine Learning: Workshop - Learning from Imbalanced Data Sets II*, 1–8.

Fisher, G. M. (1992): “The development and history of poverty thresholds,” *Social Security Bulletin*, 55.

Gordon, R. J. (2006): “The Boskin Commission Report: a retrospective one decade later,” *NBER Working Paper*, 12311.

Hoynes, H. W., M. E. Page, and A. H. Stevens (2006): “Poverty in America: Trends and Explanations,” *The Journal of Economic Perspectives*, 20, 47–68.

Jaimovich, N. and H. E. Siu (2020): “Job Polarization and Jobless Recoveries,” *The Review of Economic and Statistics*, 102, 129–147.

Jaravel, X. (2019): “The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector,” *The Quarterly Journal of Economics*, 134, 715–783.

Kaplan, G. and S. Schulhofer-Wohl (2017): “Inflation at the household level,” *Journal of Monetary Economics*, 91, 19–38.

Krueger, A. B. (2017): “Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate,” *Brookings Papers on Economic Activity*, 2017, 1–59.

Ling, C. and C. Li (1998): “Data Mining for Direct Marketing: Problems and Solutions,” *Proceedings of International Conference on Knowledge Discovery from Data*, 27, 73–79.

McDonald, J. F. and R. A. Moffitt (1980): “The Uses of Tobit Analysis,” *The Review of Economics and Statistics*, 62, 318–321.

McLaughlin, M. and M. R. Rank (2018): “Estimating the Economic Cost of Childhood Poverty in the United States,” *Social Work Research*, 42, 73–83.

Meyer, B. D., W. K. C. Mok, and J. X. Sullivan (2009): “The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences,” NBER working papers, National Bureau of Economic Research, Inc.

Meyer, B. D. and J. X. Sullivan (2012a): “Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure,” *Journal of Economic Perspectives*, 26, 111–136.

——— (2012b): “Winning the War: Poverty from the Great Society to the Great Recession,” *Brookings Papers on Economic Activity*, 43, 133–200.

——— (2013): "Consumption and Income Inequality and the Great Recession," *American Economic Review*, 103, 178–183.

——— (2019): "Annual Report on U.S. Consumption Poverty," Tech. rep.

Rank, M. R., L. M. Eppard, and H. E. Bullock (2021): *Poorly Understood: What America Gets Wrong About Poverty*, Oxford University Press.

Shaefer, H. L. and J. Rivera (2018): "Comparing Trends in Poverty and Material Hardship Over the Past Two Decades," Working paper 5-17, University of Michigan.

Sohnesen, T. and N. Stender (2016): *Is Random Forest a Superior Methodology for Predicting Poverty? An Empirical Assessment*, The World Bank.

Stock, J. and M. Watson (2015): *Introduction to Econometrics, Third Update, Global Edition*, Pearson Education Limited.

Tobin, J. (1958): "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36.

Zhao, X., B. Yu, Z. Chen, Q. Li, C. Wang, and J. Wu (2019): "Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh," *Remote Sensing*, 11.

A Machine Learning

In this section, we provide a detailed description of the random forest algorithm and its variations as they were applied in the context of our analysis.

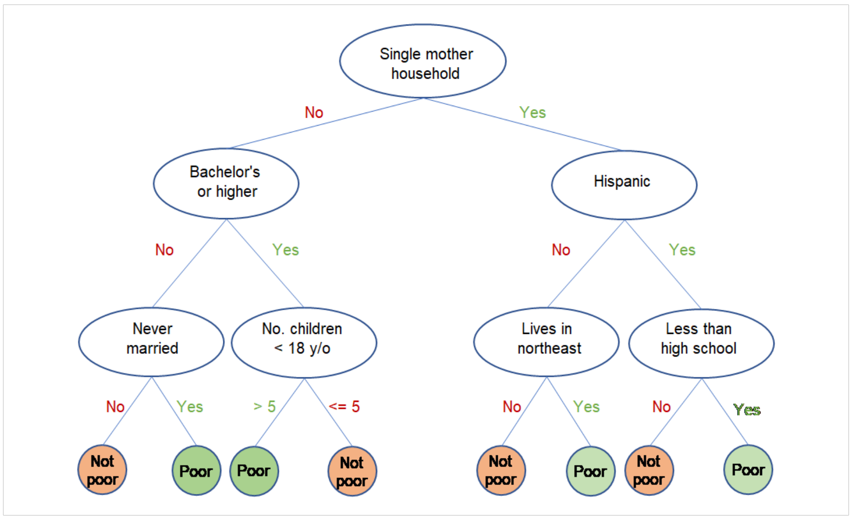
We first apply the most simple version of random forest, where the basic procedure illustrated in Figure A.1 can be summarized in the following steps:

1. Divide original sample into training (70%) and validation (30%) sets
2. For each of K trees, draw a random sample with replacement from the training set
3. Grow each classification tree based on the respective random samples
    (a) For every decision node in a given tree, randomly select a subset M from the total number of variables available
    (b) Determine which variable is used for splitting at each node based on the highest reduction in residual sum of squares
    (c) Grow each tree until a condition is met. In our case, we grow each tree until a split leads to fewer than 20 observations
4. Repeat steps 2 and 3, K times.

This method is popular among econometricians because it requires input on only two parameters: (i) the number of trees, K; and (ii) the number of variables, M, at each decision node, to be considered for the split. We rely on five-fold cross-validation in order to choose the optimal parameters. It is the case that, keeping everything else constant, $M = 5$ leads to the lowest cross-validation error. Given the choice of M, we find that after $K = 50$ trees there is no significant improvement in the cross-validation error.

Because the group of interest, poor households, represents a small fraction of our sample (roughly 11.4%), we apply a variation of traditional random forest adequate for unbalanced samples. More specifically, we use balanced random forest. The main deviation from the traditional method described above lies in Step 2. Instead, for each tree, we draw a random sample of size $2X$, where X is the number of poor households in the training set. Each strata, or group, is sampled randomly with replacement for total size X, respectively. Hence, the ratio of poor-to-non poor households is 1:1. For example, in 1984-1989 we only have 12,215 poor. This means that each tree is given a random sample of size 24,430, where the share of poor is 50%. The main drawback of this method is the loss of information regarding the majority group. This will have implications for predictive accuracy. We can expect to do a better job at predicting the poor, but also a worse job at predicting the non-poor.

Figure A.1: Random forest - A Simple Classification Tree



An alternative method considered in this analysis is gradient boosting, which is regarded as a stage-wise additive model. It combines several "weak learners" into a strong learner through an iterative process. A weak learner can be regarded as a tree that is just better at predicting than a random coin toss. The algorithm first creates a simple classification tree based on a random sample with replacement from the training set and attempts to predict whether a consumer unit is poor or not. Then at each iterative step, a new classification tree is trained on the errors (residuals) of the previous tree. Given that traditional random forest does not predict the poor very well, this method will allocate more weight to those observations with higher classification error rate (i.e. the poor), and contribute to higher predictive accuracy.

B Robustness Checks

This section presents three robustness checks for our probit analysis. We acknowledge that our baseline model in Equation 2 may suffer from omitted variable bias. Though our choice of variables in $\boldsymbol{X}$ and $\boldsymbol{H}$ was conditioned by the coverage of the Consumer Expenditure Survey, as it led to no missing observations, we were able to recover enough information on work-related characteristics of the CU head (hours worked and occupation) and whether the CU receives government support for housing. As the number of missing observations with respect to the three variables mentioned above varies both in magnitude and over time, we use three different sub-samples. Our main concern is the introduction of bias once we alter the size of the original data. If either poor or non-poor consumer units are more likely to not report information on hours worked, occupation, or whether they receive housing support from the government, the resulting sample will be biased once we drop these missing observations. More specifically, we find that the share of poor households in our sample went from 11.4% (original data) to 9.0% (hours worked), 9.1% (occupation), and 12.2% (housing support). This implies that while poor households are more likely to omit information on work-related characteristics, it is the case that non-poor households are more likely to omit whether they receive housing support from the government. Thus, we acknowledge the possibility of working with slightly biased samples in this section, though we find this exercise useful in confirming the robustness of our findings in Section 3.2.

Housing Support

Government support, if correctly targeted, can aid families facing hardship by granting access to additional resources. As a large portion of the expenditure basket for poor households is directed towards rent and food, programs such as government housing and food assistance can make a significant impact. The CES provides information on whether a household experiences lower rent due to government assistance and whether the consumer unit is part of a public housing project. We combine this information into an indicator variable, G, which takes the value of 1 if either of the above is true, and 0 otherwise, and include it in our baseline Probit model specification. This leads to the following specification:

$$\mathbb{E}(Y|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Y},G) = \mathbb{P}(Y=1|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Y},G) = \Phi(\alpha_3 + \boldsymbol{X}\boldsymbol{\beta_3} + \boldsymbol{H}\boldsymbol{\gamma_3} + \boldsymbol{T}\boldsymbol{\delta_3} + \omega_3 G). \tag{5}$$

Table B.1 presents the estimated coefficients from Equation 5, and the respective time dynamics. Overall, our main findings remain robust in that households where the household head has at least a Bachelor's degree or where both adults are married are less likely to be poor. On the other hand, families with more children, African American families, and Hispanic families are more likely to be poor, and the effects of race and ethnicity have decreased substantially over time. We also continue to see regional difference emerging in the post-GFC period where households located in the Midwest face a higher likelihood of being poor. Our findings also suggest that families that receive housing support from the government are more likely to be poor. While we cannot make any conclusion on the effectiveness of these policies on the likelihood of being poor as we do not have a panel dataset, the positive sign associated with the estimate of ω_3 can be attributed to the fact that, in proportion, poor households are more likely to receive support than non-poor CUs.

Table B.1: Probit - Housing Support

Variable	(1) 1984-1989	1990-1999	2000-2007	2008-2019	(2) Pooled sample
No. adults >64	0.115***	0.082***	0.084***	0.093***	0.088***
No. children <18	0.255***	0.255***	0.232***	0.244***	0.243***
Single mother	0.113*	0.036***	−0.101***	−0.069***	−0.016
Some college	−0.641***	−0.581***	−0.564***	−0.529***	−0.573***
Bachelor's or higher	−1.296***	−1.278***	−1.230***	−1.157***	−1.222***
Widowed	−0.683***	−0.591***	−0.557***	−0.383***	−0.513***
Separated or divorced	−0.635***	−0.488***	−0.450***	−0.309***	−0.425***
Married	−0.961***	−0.757***	−0.644***	−0.439***	−0.648***
Hispanic	0.450***	0.416***	0.400***	0.213***	0.344***
African American	0.455***	0.313***	0.246***	0.139***	0.262***
West	−0.213***	−0.194***	−0.130***	−0.065***	−0.134***
Midwest	0.027	−0.107***	−0.013	0.185***	0.024***
South	0.028	−0.030**	0.014	0.034**	0.009
Receives housing support	0.521***	0.442***	0.459***	0.245***	0.375***
Pseudo R squared		0.177			0.171
No. observations		548,740			548,740

Notes: This table shows the estimated coefficients in $\boldsymbol{\beta_3}$, $\boldsymbol{\gamma_3}$, and ω_3 from Equation 5, as well as time dynamics. ***, **, and * represent 1%, 5%, and 10% levels of significance, respectively.

Hours Worked

Although a large fraction of the poor in the U.S. are primarily adults who do not participate in the labor force and children, the "working poor" accounted for roughly 6.3 million individuals in 2019 (19% of population below the official poverty line)[7]. Given information on the usual hours worked per week by the reference person, we construct am indicator variable, F, which takes the value of 1 if the CU head usually works 35 hours per week or more, and 0 otherwise. The resulting sample, obtained by dropping observations with missing values, then consists of only consumer units where the household head is employed, which explains the reduction of the sample poverty rate from 11.4% to 9.0%. Including this information in our probit model leads to the following specification:

[7] https://www.bls.gov/opub/reports/working-poor/2019/home.htm

$$\mathbb{E}(Y|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Y},F) = \mathbb{P}(Y=1|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Y},F) = \Phi(\alpha_4 + \boldsymbol{X}\boldsymbol{\beta_4} + \boldsymbol{H}\boldsymbol{\gamma_4} + \boldsymbol{T}\boldsymbol{\delta_4} + \eta_4 F). \qquad (6)$$

Table B.2: Probit - Fulltime

	(1)				(2)
Variable	1984-1989	1990-1999	2000-2007	2008-2019	Pooled sample
No. adults >64	-0.154^{***}	-0.008	-0.034^{*}	-0.055^{***}	-0.048^{***}
No. children <18	0.264^{***}	0.265^{***}	0.245^{***}	0.248^{***}	0.251^{***}
Single mother	-0.005	-0.007	-0.058^{***}	-0.031	-0.035^{***}
Some college	-0.605^{***}	-0.542^{***}	-0.561^{***}	-0.543^{***}	-0.563^{***}
Bachelor's or higher	-1.284^{***}	-1.285^{***}	-1.225^{***}	-1.190^{***}	-1.232^{***}
Widowed	-0.936^{***}	-0.819^{***}	-0.721^{***}	-0.592^{***}	-0.725^{***}
Separated or divorced	-0.668^{***}	-0.545^{***}	-0.486^{***}	-0.437^{***}	-0.498^{***}
Married	-0.883^{***}	-0.750^{***}	-0.662^{***}	-0.571^{***}	-0.676^{***}
Hispanic	0.462^{***}	0.463^{***}	0.440^{***}	0.289^{***}	0.387^{***}
African American	0.441^{***}	0.290^{***}	0.268^{***}	0.221^{***}	0.281^{***}
West	-0.158^{***}	-0.220^{***}	-0.160^{***}	-0.095^{***}	-0.151^{***}
Midwest	-0.015	-0.089^{***}	-0.038^{**}	0.164^{***}	0.017^{*}
South	0.039	-0.053^{***}	-0.042^{**}	-0.023	-0.029^{***}
Fulltime worker	-0.678^{***}	-0.477^{***}	-0.452^{***}	-0.459^{***}	-0.491^{***}
Pseudo R squared	0.191				0.188
No. observations	442,981				442,981

Notes: This table shows the estimated coefficients in $\boldsymbol{\beta_4}$, $\boldsymbol{\gamma_4}$, and η_4 from Equation 6, as well as time dynamics. ***, **, and * represent 1%, 5%, and 10% levels of significance, respectively.

Table B.2 presents the estimated coefficients from Equation 6, and the respective time dynamics. We can observe that our findings with regards to race, ethnicity, education, region, and number of children continue to hold once we control for part-time vs. full-time employment. While the coefficients on marital status remain relatively unchanged, our results suggest that the number of elderly residing in a CU is negatively associated with the likelihood of being poor. Even after controlling for educational attainment and marital status of the reference person and other CU characteristics, we find that households headed by a full time worker are much less likely to be poor compared to a part time worker.

Occupation

Similarly to our discussion above, we incorporate information regarding the occupation of the household head. The resulting sample consists of consumer units where the household head was employed at the time of survey and has reported the most relevant occupation. The new specification reads as

$$\mathbb{E}(Y|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Y},\boldsymbol{O}) = \mathbb{P}(Y=1|\boldsymbol{X},\boldsymbol{H},\boldsymbol{Y},\boldsymbol{O}) = \Phi(\alpha_5 + \boldsymbol{X}\boldsymbol{\beta_5} + \boldsymbol{H}\boldsymbol{\gamma_5} + \boldsymbol{T}\boldsymbol{\delta_5} + \boldsymbol{O}\boldsymbol{\rho_5}), \tag{7}$$

where $\boldsymbol{O}$ represents a categorical variable for occupation. More detail can be found in Table C.1.

Table B.3 presents the estimated coefficients from Equation 7, and the respective time dynamics. First, it is important to point out that our main findings remain unchanged even after we control for the occupation of the reference person. While it is not clear whether a larger number of elderly residing in the CU is linked to a higher or lower probability of being poor, we can say that, once we control for marital status, educational attainment, occupation and the number of children, there is no clear evidence pointing to the fact that single mother households are more likely to be poor. The likelihood of being poor varies widely by occupation. Families where the reference head's job position is one of management/professional, typically associated by higher educational attainment, experience the lowest probability of being poor while households headed by an individual in lower-skilled jobs (agriculture, manufacturing, protective/personal services) are more likely to be poor.

Table B.3: Probit - Occupation

Variable	(1) 1984-1989	1990-1999	2000-2007	2008-2019	(2) Pooled sample
No. adults >64	0.007	0.059***	0.026	−0.012	0.018*
No. children <18	0.254***	0.253***	0.241***	0.247***	0.244***
Single mother	0.037	0.036	−0.024	−0.008	−0.004
Some college	−0.556***	−0.488***	−0.441***	−0.455***	−0.482***
Bachelor's or higher	−1.072***	−1.070***	−0.926***	−0.950***	−0.990***
Widowed	−0.960***	−0.817***	−0.719***	−0.519***	−0.696***
Separated or divorced	−0.733***	−0.594***	−0.524***	−0.432***	−0.526***
Married	−0.953***	−0.789***	−0.678***	−0.553***	−0.691***
Hispanic	0.356***	0.379***	0.358***	0.232***	0.308***
African American	0.355***	0.236***	0.212***	0.183***	0.225***
West	−0.169***	−0.178***	−0.144***	−0.055***	−0.123***
Midwest	−0.057**	−0.113***	−0.076***	0.150***	−0.008
South	0.038	−0.056***	−0.060***	−0.038***	−0.041**
Sales/Administrative support	0.213***	0.275***	0.336***	0.390***	0.326***
Protective/Personal/Other services	0.649***	0.611***	0.616***	0.545***	0.584***
Agriculture/Fishing/Forestry	0.786***	0.735***	0.877***	0.826***	0.820***
Construction/Mining/Mechanic	0.160***	0.260***	0.302***	0.327***	0.270***
Production line/Manufacturing	0.351***	0.514***	0.575***	0.515***	0.506***
Armed Forces	0.244***	0.089	0.098	−0.123	0.033
Pseudo R squared		0.189			0.185
No. observations		418,785			418,785

Notes: This table shows the estimated coefficients in $\boldsymbol{\beta_5}$, $\boldsymbol{\gamma_5}$, and $\boldsymbol{\rho_5}$ from Equation 7, as well as time dynamics. ***, **, and * represent 1%, 5%, and 10% levels of significance, respectively.

C Additional Tables

Table C.1: Variable description

Variable name	Type	Description
Number of children < 18	*Integer*	Indicates the number of children below the age of 18 residing in the CU at the time of the survey
Number of adults > 64	*Integer*	Indicates the number of adults over the age of 64 residing in the CU at the time of the survey
African American	*Binary*	Takes the value of 1 if the majority of the members residing the CU at the time of the survey are of African American origin, and 0 otherwise. The omitted categories are White, Asian, Others.
Hispanic	*Binary*	Takes the value of 1 if the majority of the members residing the CU at the time of the survey are of Hispanic heritage, and 0 otherwise. Ethnicity is treated differently than race in the Consumer Expenditure Survey, thus the omitted category is simply non-Hispanic.
Single mother	*Binary*	Takes value of 1 if original survey variable "Family Type" is described as "One parent, female, own children at least one $<$18", and 0 otherwise. Family type is based on the relationship of members to the reference person, whom physically occupy the address selected for the survey. Own children include biological, adopted, and step children.
South	*Binary*	Takes the value of 1 if, at the time of the survey, the consumer unit is located in the South, and 0 otherwise
West	*Binary*	Takes the value of 1 if, at the time of the survey, the consumer unit is located in the West, and 0 otherwise
Midwest	*Binary*	Takes the value of 1 if, at the time of the survey, the consumer unit is located in the Midwest, and 0 otherwise
Northeast	*Binary*	Takes the value of 1 if, at the time of the survey, the consumer unit is located in the Northeast, and 0 otherwise. This variable is the omitted category with respect to region in our Probit and Tobit regressions
Married	*Binary*	Takes value of 1 if the consumer unit reference person is married, and 0 otherwise.
Widowed	*Binary*	Takes value of 1 if the consumer unit reference person is widowed, and 0 otherwise.
Separated or divorced	*Binary*	Takes value of 1 if the consumer unit reference person is separated or divorced, and 0 otherwise.
Never married	*Binary*	Takes value of 1 if the consumer unit reference person is never married, and 0 otherwise. This variable is the omitted category with respect to marital status of the CU head in our Probit and Tobit regressions
Bachelor's degree or higher	*Binary*	Takes value of 1 if the reference person has at least a Bachelor's degree, and 0 otherwise
Some college	*Binary*	Takes value of 1 if the reference person has at least a high school degree or some college education, and 0 otherwise
Less than high school	*Binary*	Takes value of 1 if the reference person does not possess a high school degree, and 0 otherwise. This variable is the omitted category with respect to educational attainment of the CU head in our Probit and Tobit regressions
Age	*Integer*	Indicates the age of the reference person at the time of the survey
Full time worker	*Binary*	Takes the value of 1 if the reference person usually works 35 hours or more per week, and 0 otherwise.
Occupation	*Categorical*	Occupation of the reference person at the time of survey is classified into 7 main categories, following Meyer and Sullivan (2012b): Manager/Professional/Educator (base category), Administrative and technician support/Sales, Protective services/Private services/Other services, Agriculture/Fishing/Forestry/Groundskeeping, Construction/Mining/Mechanic, Production Line/Manufacturing, Armed Forces
Receives housing support	*Binary*	Takes the value of 1 if either rent costs are lower due to Federal, State, or local government payments or if the consumer unit is a public housing project, and 0 otherwise

Notes: This table presents a more detailed description of the variables in $\boldsymbol{X}$ and $\boldsymbol{H}$ considered in our analysis of the determinants of poverty, as well as additional characteristics that appear in our robustness checks in Appendix B.

PUBLICATIONS

Winning the War? New Evidence on the Measurement and the Determinants of Poverty in the United States
Working Paper No. [WP/2022/###]