AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Differences in the Distribution of Nutrition Between Households Above and Below Poverty

Rebecca Cleary, Colorado State University, rcleary@colostate.edu
Yizao Liu, The Pennsylvania State University
Andrea Carlson, Economic Research Service, USDA

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Copyright 2022 by Cleary and Liu. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Differences in the Distribution of Nutrition Between Households Above and Below Poverty*

Rebecca Cleary†, Yizao Liu‡, and Andrea Carlson§

May 16, 2022

Abstract

Nutritional inequality– that wealthier households tend to eat more healthfully than poorer ones– may contribute to health inequality via diet-related diseases, which have higher incidences among populations of lower socioeconomic status. We investigate the difference in the distribution of nutritional quality between households above and below 130% of the poverty threshold and find that across the distribution households in poverty have lower qualities, consistent with first-order stochastic dominance. Using a recently developed decomposition method applied to quantiles and a new data set linking food-at-home purchases with nutrition data, we uncover demographics, household characteristics, and food expenditures that contribute to the gap in nutritional quality. We find that the difference in nutritional quality widens as quality scores increase and that the within-group range of nutritional quality scores (10th to 90th quantile) is an order of magnitude larger than the difference between groups. The differences in demographics and household characteristics explain about 68% of the gap at the tails and about 57% at the median. Education and employment levels, specifically, contribute most to the gap while returns to those characteristics ameliorates it. Food-at-home expenditures do not appear to contribute meaningfully to the difference in nutritional quality at any point in the distribution.

JEL Codes: I1, I14, I32, I26

Keywords: nutrition, decomposition analysis, socioeconomic status, distributional analysis, Purchase to Plate Crosswalk (PPC), scanner data, IRI Consumer Network, Healthy Eating Index (HEI)

---

*The findings and conclusions in this manuscript are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. The analyses, findings, and conclusions expressed in this study also should not be attributed to Information Resources, Inc. (IRI). This project is based on research that was partially supported by the Colorado Agricultural Experiment Station with funding from the Hatch Multistate Research capacity funding program (Accession Number 7001548) from the USDA National Institute of Food and Agriculture and by the U.S. Department of Agriculture, Economic Research Service. We have no conflicts of interest to disclose.

†(contact author) Assistant Professor, Department of Agricultural and Resource Economics, Colorado State University, rcleary@colostate.edu

‡Associate Professor, Department of Agricultural Economics, Rural Sociology, and Education, The Pennsylvania State University

§Economist, Economic Research Service, USDA

1 Introduction

Diet-related, chronic diseases– like obesity, diabetes, hypertension, and high cholesterol– are linked with limited intake of nutritionally adequate foods and they disproportionately afflict households in poverty. Food assistance programs, like the Supplemental Nutrition Assistance Program (SNAP), are intended to supplement poor households' food expenditures with an aim of improving their nutrition. The 2008 Farm Bill emphasized SNAP's nutrition goals, however disparities in health and nutrition across poverty status continue to widen (Rehm et al. 2016; Seligman et al. 2010; Pappas et al. 1993; *inter alia*). Evidence of nutritional inequality is abundant (Leung and Tester 2019; Hiza et al. 2013; Leung et al. 2012; Kant and Graubard 2007; Ball et al. 2006; *inter alia*), but evidence-based policies that increase the healthfulness of diets are not (Mozaffarian et al., 2021). This is partly because the process influencing the nutritional choices of households is complex and may involve numerous factors. There is a need to better understand the gap in nutritional quality that exists between households living above and below poverty in order to improve the nutritional status and health of at-risk populations in the U.S.

Not all households are equally likely to be poor and differences in demographics, household characteristics, and other factors related to nutrition may explain, in part, the disparities in nutrition that exist. For example, households headed by women, particularly single-parent households with children, are more likely to be poor than those headed by men (Schiller, 2004; Haughton and Khandker, 2009). Households with children or a larger family size are also more likely to be poor (Provencher and Carlton, 2018). The poor tend to have fewer years of education and attend lower-quality schools; and household earning capacity is positively related with educational attainment (Rycroft, 2017). Households identifying as black are more likely to have lower incomes and fewer employment opportunities than those identifying as white (Bertrand and Mullainathan, 2004). Households living in the South and the Midwest are also more likely to be poor (Rycroft, 2017). Many of these characteristics are also associated with poorer nutrition (Rehm et al., 2016; Aggarwal et al., 2016).

To our knowledge no study to date has attempted to quantify 1) how much of the disparity in nutrition that exists between households above and below poverty is explained by demographics and household characteristics, 2) how much this disparity would be mitigated if households below poverty had similar characteristics to those above it, and 3) how the size of the disparity and factors contributing to it change over the distribution of nutritional quality. We investigate how much the difference in the composition of households contributes to the gap in nutrition across poverty status. Comparisons of means may mask differences in the extremes of the distribution, where health outcomes of nutrition are more pronounced (Silbersdorff et al., 2018). Given that households with poorer nutrition are more at-risk for diet-related diseases than those with better nutrition (Onvani et al., 2017; Harmon et al., 2015), a 'beyond the mean' approach in which the entire empirical distribution is considered allows us to make comparisons based on the level of nutritional quality. We first use Unconditional Quantile Regression (UQR; Firpo et al., 2009) to estimate the relationship between demographics and other household factors and nutritional quality separately for households above and below poverty. We then investigate the difference in nutritional quality between households above and below poverty at every fifth quantile of the distribution of nutritional quality using a decomposition method recently developed by Firpo et al. (2018), hereafter FFL decomposition, based on recentered influence functions (RIF). We decompose the differences at each quantile into two components, the difference in: 1) the levels of determinants (composition or policy effect) and, 2) the relationship of those determinants to nutritional quality (structure effect). Using USDA's recently released Purchase-to-Plate Crosswalk (PPC), which combines food purchase data with its nutritional equivalents (Carlson et al., 2019), we are able to uncover differences in nutritional quality via the foods households purchase throughout the year, rather than on a few days of dietary recall data. We estimate the Healthy Eating Index 2010 (HEI) for each household-month for 2013. The HEI is a widely used measure of adherence to the *Dietary Guidelines for Americans 2010* (DGA)[1] and is also related with diet-related disease prevalence, as individuals at the upper tail have a reduced risk of morbidities and mortality (Onvani et al., 2017). The HEI is an improvement over other measures of nutrition that have recently been used with household scanner data because it includes nutrients from all retail food purchases and is a valid and reliable measure of the quality of a diet (Guenther et al., 2014).

Studies at the mean support a -4 to 7 point difference in HEI scores for households of varying levels of poverty (Hiza et al., 2013), where the negative value indicates that the poorer sample had, on average, higher HEI scores. Leung and Tester (2019) find a difference of about 2.2 HEI points between food-secure and food-insecure households. Differences are larger when comparing households participating in SNAP. Singleton et al. (2020) find a difference of about 5.04 HEI points between households eligible and ineligible for SNAP using two days of dietary recall and Mancino et al. (2018) find a difference of about 5.24 HEI points using a week of food acquisitions. Mean differences in HEI scores point to discrepancies, but they may also mask differences in the underlying distributions. When clinical recommendations are different along the distribution, as in the case of nutritional quality, it is more appropriate to compare distributions than means. Studies that find reduced disease incidence and mortality from adherence to the DGA generally do so by comparing the upper and lower quantiles, which is a difference of about 40 HEI points (Harmon et al., 2015). So it is not clear that the analyses done at the mean support a meaningful difference in HEI scores from a clinical perspective. By comparing the distributions of nutritional quality for households above and below

[1]https://www.dietaryguidelines.gov/about-dietary-guidelines/previous-editions/2010-dietary-guidelines

poverty, we can contextualize the difference across poverty status against the difference between the upper and lower quantiles of each subpopulation.

An increasing number of papers in the health economics literature highlight the need to look beyond the mean and use distributional methods to uncover associations with health or health outcomes (Silbersdorff et al., 2018; Bilger et al., 2017; Carrieri and Jones, 2017; Heckley et al., 2016; Makdissi and Yazbeck, 2014; Duclos and Échevin, 2011). Silbersdorff et al. (2018) use structured additive distributional regression and find that the poor face greater health risks at the lower tail of the health distribution. Bilger et al. (2017) use distribution-sensitive indices to understand how the depth and severity of obesity vary across socio-economic status and find that obese poor weigh significantly more than obese non-poor, which increases their risk for adverse health outcomes and is masked by analyses of categorical measures at the mean. The relationship of socio-economic factors and body mass index (BMI) has also been studied across the distribution of BMI (Bann et al., 2020; Bonanno et al., 2018). Carrieri and Jones (2017) use UQR and uncover a non-linear relationship between income and health bio-markers that differs substantially across gender. They find that the relationship between ferritin, an indicator of sufficient nutrition, and income is positive, but small, and becomes larger toward the upper middle of the distribution of ferritin. They do not find a statistically significant relationship between bio-markers for diabetes and income until the 80th quantile; the 95th being the clinical cutoff for a diabetes diagnosis. We contribute to this literature by bringing a distributional approach to the estimation of differences in nutritional quality across poverty status.

In the health economics literature, FFL decomposition has been used previously to investigate contributing factors to the inequality of socio-emotional skills between genders (Attanasio et al., 2020) and the mental health gap between genders (Churchill et al., 2020). Other decomposition techniques have been used to uncover inequality in health within and between countries, differences in short- and long-run measures of health, and inequalities in the presence of certain biomarkers, among others (Davillas and Jones 2020; Carrieri and Jones 2018; Jones and Nicolás 2004; Pradhan et al. 2003; *inter alia*).

By using retail food purchase data, we are able to link nutrition with actual expenditures, which has not been possible with other national surveys or clinical data. Our findings extend the understanding of how the cost of a healthy diet contributes to the poverty gap in nutrition. Analyses of food expenditures and nutritious diets are often complicated because they require the combining of data from different sources. Many rely on food items consumed at the individual or household level, but calculate cost based on national average food prices. Because lower-income households are able to pay a lower price per unit for the same product than higher-income shoppers even within the same store, analyses relying on national average prices or expenditures to proxy for quantity may not be accurate (Broda et al., 2009). Our data set includes the price paid and nutrients for each item, allowing us to have a much better understanding of diet costs at the household level. We find that households at the upper quantile of nutrition quality spend about $20/month more on food-at-home per person than those in the lower quantile for both households above and below poverty; however, the difference in expenditures across poverty status contributes less than 5% to the gap in nutrition between the two groups.

There are very few studies that consider the entire distribution of nutritional quality when comparing different populations. For example, Variyam et al. (2002) use a conditional quantile regression (CQR) approach to investigate how macronutrient intake varies across demographic factors, but, because CQR yields estimates conditional on the value and presence of covariates, they cannot compare the distributions of demographic subpopulations. By using the unconditional distribution, we are better able to understand the relationship of demographics and other household factors with nutrition, unconditional on the presence or values of other covariates. Further, by using the unconditional distribution for the decomposition analysis, we can compare the relationship between demographics and household factors, like food expenditures, to the gap in nutrition across poverty status.

We find that at the lower values of the HEI distribution, the difference in nutritional quality for households above and below poverty is slight, a little more than 1 percentage point and it is much more marked at the higher quantiles of nutritional quality, about 3-4 percentage points. This contrasts with findings for obesity, from which the poor suffer more severely (Bilger et al., 2017) and also suggests that focusing on ameliorating the nutritional-quality gap between households above and below poverty will not result in a meaningful increase in nutritional quality for poorer households at the lower end of the nutritional quality distribution, where households are most at-risk for increased morbidities and mortality. We find that differences in the composition of each group contributes about 68% to the gap in HEI scores at both tails and about 57% around the median. Differences in food expenditures contribute less than 5% for most of the distribution. In contrast, differences in education and employment contribute from about 38% at the lower tail to about 50% at the upper tail. We also find that below poverty households at the upper end experience greater returns to nutritional quality from post-graduate education than households above poverty and this slightly mitigates the gap in nutritional quality at the upper tail.

This paper proceeds as follows. First, we discuss the empirical model and how we use the FFL decomposition method to understand which factors contribute most to the gap in nutrition. Then, we describe the household scanner data, the PPC, and the elements we used to estimate the nutritional quality of household purchases. We compare the cumulative distribution functions for evidence of differences before proceeding with the main analysis. We follow with a discussion of our results and conclusions.

2 Empirical model and methods

Our objective is to understand how household attributes contribute differently to the nutrition of households above and below poverty. To that end, we begin by partitioning households into those above, $T = 1$, and below, $T = 0$, poverty and then consider a simple reduced-form equation that includes socioeconomic status and demographic variables,

$$h_{iT} = \alpha_T + Z_{iT}\beta_T + D_{sT}\delta_{sT} + D_{mT}\delta_{mT} + \epsilon_{iT}, \text{ for } T = 1, 0 \tag{1}$$

where $h_{iT} \in H$ represents nutritional quality for households above and below poverty. Z_{iT} is a vector of household socioeconomic status and demographic characteristics that have been found to be important to explain nutrition or diet quality. Eq. 1 can be thought of as the reduced-form version of the household production model in which households choose foods to purchase within the context of their demographics, food preferences, and resources (Singleton et al., 2020; National Research Council and the Institute of Medicine, 2013). In their study of HEI over time, Rehm et al. (2016) found that age, sex, race and ethnicity, education level, and family income were important demographics for nutrition and identified diet quality discrepancies by race and ethnicity, education, and income level. Other studies have included education, race, age, employment, occupation, stressors, children, household size, sex, and martial status to explain HEI (e.g., Allcott et al., 2019; Leung and Tester, 2019; Smith et al., 2019; Hiza et al., 2013). We include education, race, age and presence of children, employment, household size, ethnicity, and marital status. Given that we are partitioning our sample by poverty status, we do not include income, however this method allows the marginal effect of all variables to vary with poverty status. We also include fixed effects to control for some unobservables: month fixed effects, D_{mT}, capture the variation of household nutritional quality over the course of the year and state fixed effects control for spatial differences in nutrition. D_{sT}. α_T, β_T, δ_{sT}, and δ_{mT} are parameters to be estimated and ϵ_{iT} are the stochastic error terms. For compactness, let Z_{iT}, D_{mT}, and $D_{sT} \in X_T$.

2.1 Unconditional quantile regression

To understand how the relationships given by eq. 1 vary across the distribution of nutritional quality, we use the UQR estimation procedure based on RIFs developed by FFL. UQR offers our analysis several distinct advantages. UQR is able to capture distributional impacts that may be masked by the mean because it is evaluated marginally over the distribution, which allows us to capture differences at the tails of the distribution, which are of clinical importance whereas the mean is not. Also, as mentioned previously to compare two distributions, CQR falls short under general assumptions as it conditions on the presence and values of covariates and not all households are equally likely to be poor.

The RIF unconditional quantile summarizes the impact of an individual observation on each quantile. It is essentially a linear approximation to the non-linear quantile function and measures how a change in the underlying distribution impacts each quantile value. The RIF of the τth quantile is

$$RIF(h_{iT}; q_\tau, F) = q_\tau + IF(h_{iT}; q_\tau, F) = q_\tau + \frac{\tau - \mathbb{I}\{h_i < q_\tau\}}{f_h(q_\tau)} = c_{1,\tau} \cdot \mathbb{I}\{h_i < q_\tau\} + c_{2,\tau}, \tag{2}$$

where $c_{1,\tau} = {}^1/_{f_h(q_\tau)}$, $c_{2,\tau} = q_\tau - c_{1,\tau} \cdot (1-\tau)$, and $f_h(q_\tau)$ is the density of H evaluated at q_τ. Therefore, the expectation of the RIF can be written as

$$\mathbb{E}[RIF(h_i; q_\tau)|X_T] = c_{1,\tau} \cdot \Pr[H > q_\tau|X_T] + c_{2,\tau}. \tag{3}$$

We specify the RIF regression to estimate eq. 1 as

$$\mathbb{E}[RIF(h_i; q_\tau)|Z_i, D_s, D_m, T] = \alpha_T + Z_{iT}\beta_T + D_{sT}\delta_{sT} + D_{mT}\delta_{mT} + \epsilon_{iT}, \text{ for } T = 1, 0 \tag{4}$$

The parameters from the RIF regression are estimates of unconditional marginal effects; the unconditional property of the UQR implies that the coefficients of the RIF have a similar interpretation to those from an OLS regression (Firpo et al., 2009), but applicable to a particular quantile.

2.2 Two-stage decomposition

To compare and decompose the gaps in the quantiles of nutritional quality between households above and below poverty, taking into account any linear and nonlinear differences in the joint distribution across those groups, we use RIF regressions, which yield the partial effect of a small location shift in the distribution of covariates on the quantiles but are not appropriate for analyzing the impact of a large change, such as the difference of the distribution of nutritional quality for two groups (Firpo et al., 2018). A useful method to decompose larger differences would be the Oaxaca-Blinder decomposition, which provides a way of decomposing differences into a structural effect and a composition effect and also allows for the decomposition of those effects into the contribution of each covariate. However, the Oaxaca-Blinder decomposition can only be applied at the mean and would not allow us to uncover effects across the distribution of nutritional quality. Firpo et al. (2018) develop the use of RIF regressions combined with the re-weighting strategy proposed by DiNardo et al. (1996) to decompose quantiles, hereafter referred to as FFL decomposition. The first stage consists of estimating weighting functions $\omega_1(T)$, $\omega_0(T)$, and $\omega_C(T)$, which express the probability of being in each group. The first two weights can be computed directly from the empirical distributions as $\hat{\omega}_1(T) = \frac{T}{\hat{p}}$ and $\hat{\omega}_0(T) = \frac{1-T}{1-\hat{p}}$, where $\hat{p}$=$\frac{1}{N}\sum_{i=1}^{N} T_i$. The counterfactual distribution is not directly observed in the data; it is constructed using the observed characteristics of households below poverty and multiplying this by a weighting factor so that it resembles the distribution of households above poverty. We estimate the weights of the counterfactual as $\hat{\omega}_C(T, X) = \frac{1-T}{\hat{p}}\left(\frac{\hat{p}(X)}{1-\hat{p}(X)}\right)$ where $p(\cdot)$ is a logit estimator of the conditional probability of being above poverty given X. We compute the quantiles $q_{\tau 1}$, $q_{\tau 0}$, $q_{\tau C}$ as follows

$$\hat{q}_{\tau t} = \operatorname{argmin}_q \sum_{i=1}^{N} \hat{\omega}_t(T_i) \cdot \rho_\tau(h_{it} - q), t = 1, 0, \tag{5}$$

$$\hat{q}_{\tau C} = \operatorname{argmin}_q \sum_{i=1}^{N} \hat{\omega}_C(T_i, X_i) \cdot \rho_\tau(h_{it} - q), \tag{6}$$

where $\rho_\tau(h_i - q)$ is the check function of Koenker and Bassett Jr (1978).

The estimators for the differences in the quantiles are given by

$$\hat{\triangle}_S^{q_\tau} = \hat{q}_{\tau 1} - \hat{q}_{\tau C} \tag{7}$$

and

$$\hat{\triangle}_X^{q_\tau} = \hat{q}_{\tau C} - \hat{q}_{\tau 0}. \tag{8}$$

In the second stage, we estimate linear RIF-regressions. First, the RIF is computed for each observation by substituting the sample estimate of the quantile, $\hat{q}_\tau$, and estimating the density at the sample quantile, $\hat{f}(\hat{q}_\tau)$. The RIF regressions are estimated by replacing the dependent variable, h_{it}, with the estimated value of $RIF(h_i|q_{\tau t}, F)$. The regression coefficients are calculated as

$$\hat{\gamma}_t^{q_\tau} = \left(\sum_{i=1}^{N} \hat{\omega}_t(T_i) X_i X_i'\right)^{-1} \cdot \sum_{i=1}^{N} \hat{\omega}_t(T_i) X_i R\hat{I}F(h_i|q_{\tau t}, F_t), \quad t = 1, 0 \tag{9}$$

and

$$\hat{\gamma}_C^{q_\tau} = \left(\sum_{i=1}^{N} \hat{\omega}_C(T_i, X_i) X_i X_i'\right)^{-1} \cdot \sum_{i=1}^{N} \hat{\omega}_C(T_i, X_i) X_i R\hat{I}F(h_i|q_{\tau C}, F_C). \tag{10}$$

The structure effect is estimated as [2]

[2]Note that this is the difference in the coefficients between group 1 and the counterfactual when group 0 is reweighted to have the same distribution of covariates as group 1; not the traditional Oaxaca-Blinder difference in coefficients. That is, $\hat{\triangle}_P^{q_\tau}$ only reflects the differences in marginal nutrition returns from covariates and is not contaminated by differences in the distribution of covariates among the two groups (Firpo et al., 2018).

$$\hat{\triangle}_P^{q_\tau} = \mathbb{E}[X, T=1]'(\hat{\gamma}_1^{q\tau} - \hat{\gamma}_C^{q\tau}), \tag{11}$$

and the composition effect is estimated as

$$\triangle_X^{q_\tau} = (\mathbb{E}[X|T=1] - \mathbb{E}[X|T=0])' \, \hat{\gamma}_0^{q_\tau} + \hat{R_X}^{q_\tau}, \tag{12}$$

where

$$\hat{R}_X^{q_\tau} = \mathbb{E}[X|T=1]' \, (\hat{\gamma}_C^{q\tau} - \hat{\gamma}_0^{q\tau}) \tag{13}$$

is an estimate of the specification error.[3]

We then repeat the FFL decomposition using the households below poverty and the opposite counterfactual nutritional quality structure as the reference, which allows us to decompose the differences in the quantiles and to calculate the error associated with reweighting, $\hat{R_S}^{q_\tau}$, which should go to zero in large samples (Firpo et al., 2018). When using the Oaxaca-Blinder decomposition at the mean, many studies report the portion "explained" by endowments of characteristics and leave the remainder of the gap in an "unexplained" portion. Using the FFL approach and repeating the reweighting process allows us to decompose the "explained" portion into the composition effect and the specification error and the "unexplained" portion into the structure effect and the error from reweighting; which also allows us to avoid the traditional Oaxaca-Blinder problem of results varying with the choice reference group.

3 Data

The IRI Consumer Network Panel (IRI-CNP) database provides the core data for our analysis; it records all food-at-home (FAH) purchases (prices and quantities) at the purchase occasion level and annual demographic information for a nationally representative sample of about 100,000 U.S. households for 2013 (Sweitzer et al., 2017; Muth et al., 2016). The households we use belong to the "Static Panel," which is a panel of households meeting specific criteria for purchase recordings and which have assigned sample weights to result in nationally representative consumer purchases and is about 60,000 households across the year. The

[3] FFL note that using a linear specification for the RIF-regression instead of a more general functional form changes the interpretation of the specification error by adding an error component linked to the fact that a potentially incorrect specification may be used. Even with this potential error, FFL recommend using the linear specification for three reasons: 1) FFL's procedure only gives a first-order approximations of large changes in X, so regardless of functional form, there will be a specification error; 2) the linear specification does not affect the overall estimates of the structure and composition effects that are obtained using the re-weighting procedure; 3) the linear specification provides a straightforward interpretation of the decomposition (comparable to the Oaxaca-Blinder decomposition interpretation). When using a linear specification, FFL also suggest to measure the size of the specification error to make sure that the FFL approach provides an accurate enough approximation for the specific research question.

IRI-CNP data face similar limitations as other household scanner data. First, they do not include food purchased to eat away from home, for example in restaurants or schools.[4] Second, most households only record purchases of packaged items with bar-codes and not products sold with random weight such as bulk produce or grains.[5] Due to its comprehensiveness the IRI-CNP has been found appropriate for use in health policy research (Zhen et al., 2019).

We measure the nutritional quality of retail food purchases using the HEI, which is a validated measure of diet quality, independent of quantity, that can be used to assess adherence to the DGA (Guenther et al., 2014).[6] The HEI uses a 100 point scale divided between 12 components– nine adequacy and three moderation components (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2010); higher scores reflect greater nutritional quality.[7] Although some nutrition information is included in the IRI-CNP, it is not sufficient to estimate HEI scores for the foods purchased because it does not contain the food pattern equivalencies– which converts food quantities into their edible portions (i.e., the non-edible peels, skins, bones, and shells are removed). USDA's newly created Purchase-to-Plate Crosswalk (PPC) links the USDA data with the IRI scanner data and includes conversion factors so that the food pattern equivalencies can be properly applied to purchases (Carlson et al., 2019). Previous studies using scanner data to investigate nutrition had to develop their own measures of healthfulness, which were based on the purchases of a subset of items (e.g., fruits and vegetables). We are able to overcome this limitation by linking the PPC with scanner data to calculate a valid and reliable measure of nutritional quality that incorporates all retail food purchases. HEI, as a density measure, can be calculated from any level of aggregation of PPC-linked purchase occasions. The trade-off we must make is between the level of granularity and an accurate reflection of the nutritional quality of households' retail food purchases. We aggregate purchase occasions by month for two main reasons: 1) monthly purchases more accurately reflect home inventories and 2) many poor households (e.g., those on SNAP) make a large shopping trip once a month. We calculate HEI scores for each household-month in the 2013 Static Panel according the algorithms designed by the National Cancer Institute (Guenther et al., 2013).

We classify households as above poverty (AP) or below poverty (BP) based on 130% of the US Poverty

[4]For all income groups, the share of macronutrients consumed away from home is about the same as the share of calories consumed away from home, so grocery purchases are not a systematically biased measure of nutritional quality (Allcott et al., 2019).

[5]Allcott et al. (2019) use a subsample of households which recorded prices and weights of non-packaged items and find that 60% of produce calories are from bar-coded items for similar scanner data. They do not find evidence that this percentage varies with income and conclude that given produce is a small share of overall retail food purchases and that packaged produce is a 'significant and reasonably representative portion' of overall produce purchases that estimating nutritional relationships with scanner data is valid.

[6]Given that our IRI data are from 2013, we opt to calculate the HEI based on the DGA from 2010 rather than that corresponding to the DGA 2015 because it is the contemporaneous standard.

[7]We are not measuring the healthfulness of foods eaten or in the home inventory; therefore we follow others (e.g., Allcott et al., 2019) and refer to an HEI calculated from retail food purchases as a measure of nutrition instead of "diet quality."

Thresholds for 2013 (U.S. Census Bureau 2013), which are determined based on income and household size. 130% of the poverty thresholds is also the cutoff to determine initial eligibility for SNAP and other federal means-tested food assistance programs. We drop all observations for which total food-at-home expenditures were less than $20 per month or which exceeded $1000 per person per month (a total of 16,575 observations, less than 2.3% of the sample). We also drop 110 household-months which were estimated to have an HEI score of 0; which leaves us with 78,599 observations for BP and 634,494 observations for AP.

Weighted means and standard deviations are presented in table 1 for AP and BP. Notably, even relatively higher nutritional qualities in the U.S. still have low adherence to DGA and are in need of improvement. The weighted mean HEI for BP is 41.62; for AP it is 44.47; on average, BPs' diet qualities are about 2.9 points lower than AP.[8] These estimates of HEI are lower than others' using different data sets; however, the difference between BP and AP HEIs is largely inline with others (Hiza et al., 2013).[9]

Figure 1 presents kernel density estimates of HEI for AP and BP using the Epanechnikov kernel and a bandwidth of 1.14. FFL recommend examining the density of nutrition for unusual features that would hinder the estimation of the RIF at the quantiles of interest. Issues to look for include peaks associated with heaping, which can result from HEI components being top- and bottom-coded (Firpo et al., 2018; Fortin et al., 2011).[10] The density of HEI is quite smooth for both AP and BP. The RIF framework may not perform well at the extreme tails of the distribution and therefore we restrict our analysis to the 5th through the 95th quantiles; which correspond to HEI scores of 25.43 to 67.34 for AP households and 24.36 to 63.77 for BP.

The IRI-CNP data has five categories of education (measured by the highest level attained: primary school, some high school, high school diploma, some college, college diploma, and post-graduate work), four categories of race of the household head (White, Black, Asian, other), eight categories of ages and presence of children which we aggregate to four (under 6, between 6 & 13, between 13 & 18, no children), three categories of employment of the household head (employed less than 35 hours/week, employed more than 35 hours/week, homemaker/student), eight values for household size (with the eighth category serving to indicate the presence of at least eight members), two ethnicities of the household head (Hispanic and non-Hispanic), and four categories of marital status (married, widowed, divorced/separated, and single).[11]

[8] To contextualize, having at least 1.1 cups of vegetables, 0.8 cup of fruit, or 2.5 oz of protein per 1,000 calories each contributes 5 points to the total HEI-2010 score.

[9] Hiza et al. (2013) calculate average HEI scores based on poverty status using one day of dietary intake; their calculations of adult scores range from 51 (for households between 130% and 299% of the poverty thresholds) to 62 (for households above 500% of the poverty thresholds). They calculate that households below 130% of the poverty threshold have an average HEI score of 55.

[10] Censoring of observations within the distribution can pose a challenge to RIF estimation if they signify an unusually high value of the density at a particular quantile of interest, which can bias the estimation of the influence function (Firpo et al., 2018; Fortin et al., 2011). Many of the distributions of the HEI-2010 components are bi-modal, with heaping at the lowest and highest values; this heaping encompasses significant portions of the distributions and therefore, the techniques used in this paper are inappropriate to measure the discrepancy of component purchases.

[11] 'Age and presence of children' is our only categorical variable that is not mutually exclusive. A household can belong

There are 10 occupation categories which we aggregate to three (white collar, blue collar, and no or other occupation). Other variables available are discussed in Muth et al. (2016). In the IRI-CNP, the household head is the primary shopper. The household head can be either or both adult heads of the household.[12] Given that the number of household heads may be arbitrarily determined by the household, we recode education, employment, and occupation variables to indicate only the highest category to which a household belongs. On average, more BP indicate that primary school, some high school, and some college are the highest levels of education attained whereas more AP indicate that a college diploma or post-graduate work are the highest education attained. On average, more BP identify as Black or other race, have children of any age, are employed part-time, have larger household sizes, are Hispanic, and are married or divorced. In contrast, on average, more AP identify as Asian, are employed full-time and are married.

4 Results and discussion

4.1 Results from the unconditional quantile regressions

The parameter estimates from the UQR estimation of eq.4 for AP and BP are presented in figures 2 and 3, respectfully. In the following, our estimated coefficients can be interpreted as the marginal change in HEI points associated with the variable under consideration. In general, our results support the findings of Aggarwal et al. (2016) who find that lower socioeconomic groups are able to achieve diets comparable in quality to those of higher socioeconomic status groups. We conduct equivalence tests of the OLS and UQR estimates at every quantile, which are available in the appendix; we reject equivalence between the OLS and UQR estimates at a quantile with a 5% confidence level if the OLS estimate lies outside of the 95% confidence interval of the UQR following Hoenig and Heisey (2001). [13]

4.1.1 Households above 130% of the poverty threshold (AP)

FAH-EPP are most positively associated with nutrition at the lower tail; we estimate that a dollar increase in FAH-EPP is associated with about a 0.01-point increase in HEI for the 5th - 15th quantiles. Between those quantiles, the average FAH-EPP is about \$115, translating into an association of 1.15 HEI points at the lower quantiles. In the middle of the distribution, the relationship is slightly weaker– about 0.007, with

to more than one of the categories indicating that a child is present in the household; however it cannot belong to both the excluded group (no children present) and one of the included groups.

[12] Values are specifically coded for male and female household heads. While we recognize that some households may have heads that do not identify as either male or female or that multiple heads of the same gender may exist, to our best understanding, the data collection process used does not allow for information consistent with these types of households to be accurately recorded.

[13] We also conduct equivalence tests of the OLS and UQR estimates at every quantile, which are available in the appendix; we reject equivalence between the OLS and UQR estimates at a quantile with a 5% confidence level if the OLS estimate lies outside of the 95% confidence interval of the UQR following Hoenig and Heisey (2001).

an average FAH-EPP of about $118, translating into an association of 0.83 HEI points. Similarly for the upper tail, for which we fail to reject equivalence with the OLS estimate of 0.0084. These findings are inline with Rehm et al. (2015) who also find evidence of a positive association between diet costs and HEI scores across the distribution of diet costs, using a national food price database to estimate costs.

Education has the largest estimated association with the nutrition of AP and is greatest for households above the median. The OLS estimates for education range from -0.70 for some high school to 4.15 for post-graduate degrees and represent the marginal effect of education on the conditional mean of household HEI (where high school degree is the reference category). These estimates mask the effect of education across the distribution. For households in the lower half of the distribution of dietary quality, education contributes much less to nutrition than for households in the upper half. For example, the contribution of a college degree to AP HEI is 0.77 points for households in the 5th quantile while its contribution is about 2.46 points for households in the 75th quantile. Households in the 5th quantile of HEI have a 1.15-point increase in HEI score associated with a post-graduate degree while households in the 75th quantile experience about a 5.7-point increase. Investigating changes in the means of a modified HEI score, Allcott et al. (2019) suggest that improved educational opportunities could play a role in reducing nutritional inequality. Our findings, looking across the unconditional distribution of HEI, suggest that policies based on a link between education and dietary quality may benefit those households with the highest dietary qualities more than those with lower scores. We fail to reject equivalence between the OLS and UQR estimates of primary school and some high school for almost all quantiles; their associations with HEI remain relatively constant across the distribution.

The indicator for identifying as Black has negative coefficient estimates for all quantiles which are statistically significant at the 1 percent level. The OLS estimated coefficient is -0.84, which seems to reflect the upper tail of the distribution reasonably well. However, at the median the estimated coefficient is -1.19, which is about 41% larger than the OLS coefficient and at the lower tail it is about 32% lower. Given the size of the coefficients, reducing racial disparities may have a positive impact on Black households' nutrition, including those households with scores at or below the median, which are most at risk for diet-related diseases. The indicator for households that identify as Asian is negative and statistically significant for the lowest quantiles; it is not statistically significant at the 1 percent level for the middle quantiles. It becomes positive at higher quantiles, but lacks statistical significance at the upper tail. The OLS estimate may be a relatively good measure for the association between 'other race' and dietary quality; we fail to reject equivalence with the OLS estimate for all quantiles. White collar occupations (versus no collar) are positively associated with nutrition quality, and more so at the upper quantiles. Blue collar occupations, instead, are negatively associated with nutrition quality, and more so until the median where it levels off until the 70th quantile and then slightly increases. Younger children (under 6) are positively associated with nutrition across the distribution, although more so at the middle than the tails of the distribution. At the upper and lower tails, an additional child under 6 is associated with about a 1-point increase in HEI. At the median, it is associated with a 1.72-point increase. Children between 6 and 13 are also positively associated with nutrition, and more so at the lower tail and median than the upper tail. In contrast, teenagers are not associated with a statistically significant change in nutrition for households below the 35th quantile and beyond that, there is a negative association. This adds granularity to others' findings that children have a negative effect on dietary quality (Tiffin and Arnoult, 2010).

We find that being married (versus single) is associated with slightly higher increases in nutrition for households below the median (about 0.47 points) and lower values for those at the upper tails. The nutrition protection that marriage affords is underestimated by OLS for most of the distribution and is overstated for households at the upper tail. For divorced and widowed households, we instead find an increasingly negative relationship with nutrition across the distribution, with households at the upper tail having the most losses in nutrition associated with the end of a marriage. Versus the reference group of not employed, we find that being employed part or full time has similar negative associations and across most quantiles, we fail to reject equivalence with the OLS estimates. The OLS estimates of -0.29 and -0.33 for part and full time, respectively, are good approximations of the relationship between employment and nutrition across the distribution. Many papers have documented the negative relationship of nutrition quality and household size. We find that at the lower end of the distribution (below the 20th quantile), household size has a positive relationship with HEI scores of about 0.2. Beyond the 20th quantile, the relationship becomes increasingly negative, reaching its largest magnitude at the 80th quantile (-0.68 HEI points). Household size is not negatively associated with nutrition for those households with the worst nutrition.

4.1.2 Households below 130% of the poverty threshold (BP)

FAH-EPP has the largest associations with HEI at the lower quantiles– about 0.009 HEI points for an additional dollar of FAH-EPP. BP at these quantiles spend about \$100 FAH-EPP; at that level of spending, the association is about 0.9 HEI points. The association lessens across the middle: it is about 0.005 HEI points for an additional dollar of FAH-EPP, which is about 0.53 HEI points when spending \$105. It increases slightly at the upper tail: about 0.007 HEI points at the margin, which is about 0.77 HEI points when spending \$110. These are not large values, nor are they directly comparable with the estimates for the AP; to compare and contextualize the contribution of FAH-EPP to the gap in the HEI distributions we turn to our decomposition results in the next section.

Similarly to AP, education has the largest association with BP nutrition. Less than a high school diploma is generally negatively associated with nutrition across the distribution, while levels beyond high school have a general, positive association. The association between education and nutrition is lowest at the lower tail of the distribution, where households are most at risk, for all levels of education (excluding primary school, for which we fail to reject equivalence with the OLS estimate across most quantiles). At the lower tail, the associations with some college or a college diploma are not statistically significant; at the upper tail, these education levels are associated with a 0.89 and 2.93-point increase in HEI, respectively. A post-graduate degree is associated with higher HEI scores, which increase across the distribution: at the 10th quantile the estimate is 1.43 and at the 90th it is 7.58.

BP that identify as Black are associated with lower HEIs across the distribution and we fail to reject equality with the OLS estimate of -0.78 for most quantiles. We do not find statistically significant associations between identifying as Asian or other for most of the distribution.[14] Reducing racial disparities may be associated with a slight increase in nutrition among Black BP households. We also do not find significant variation in the relationship of occupation with nutrition across the distribution: we fail to reject equality with the OLS estimate of 0.31 for white collar jobs (vs no collar). There is some slight variation for blue collar jobs: at the 25th quantile, we do not find a statistically significant relationship, but at the 70th quantile we find that blue collar jobs are associated with about a point lower HEIs.[15]

Children under 6 contribute positively to HEI; and this association is strongest for higher levels of nutrition: at the 15th quantile the estimate is 1.17 and at the 85th it is 2.30. We fail to reject equivalence with the OLS estimate for children between 6 and 13: they are associated with about a 0.41 point increase in HEI across the distribution. Teenagers contribute positively to HEIs below the median and negatively to those above: about 0.55 points at the 10th quantile and about -1.31 points at the 90th. The protective effects of marriage (vs being single) are similar for divorcees and widow(er)s at the lower end of the distribution and they are small. The estimates get increasingly negative across the distribution: for healthier BP marriage, widowhood, and divorce are associated with lower HEIs. Part-time employment is generally positively associated with HEIs, and full-time more so. We fail to reject equivalence with the OLS estimates of 0.44 and 0.76 for part- and full-time work, respectively. Household size has a null association with nutrition until about the 40th quantile and then it has an increasingly negative association with HEI.

4.2 Decompositions

Table 2 shows the results of decomposing the total difference in HEI scores into the structural and composition contributions to the gap in nutrition. Recall that the gap in nutrition between AP and BP is quite small

[14] Note that the confidence intervals are quite large and do include meaningful values.

[15] Although note the large confidence intervals: at the 70th quantile, we cannot reject equivalence that the association is between -1.45 and -0.67.

at the lower end of the distribution and widens as nutrition improves. We find that at Q5 the difference is about 1.1 HEI points, at the median it is 3.2, and at Q95 it is 3.4. It reaches a maximum difference of 4.0 at Q80. How large is a meaningful difference in HEI scores is an open question and, as noted above, there is evidence that our mean HEI differences are slightly underestimated compared to those using data on all purchases or diet (by about 2.14 to 2.34 HEI points). Many studies finding a link between increased morbidities or mortality and HEI compare the extremes of the distribution which often represent a difference of about 40 points (see e.g., Harmon et al., 2015). One way that Guenther et al. (2014) validate the HEI as a measure of diet quality is by finding differences in the HEIs between groups known to have different diet qualities: these differences range from 2.9 to 10.7. Kirkpatrick et al. (2018) claim that a difference between independent groups of about 5 to 6 points may be meaningful. With this in mind, it is not clear that policies aimed to ameliorate the gap in nutrition between AP and BP would benefit BP at the lower tail of nutrition, where households are most at-risk for increased morbidities or mortality.

We find that differences in the demographic and household-factor composition of each group contributes about 68% to the gap in HEI scores at both tails and they contribute about 57% around the median. This is inline with recent findings that demographics explain about 72% of the disparity in diet quality at the mean between food assistance program participant and income-ineligible non-participants (Singleton et al., 2020), and suggests that their mean estimate may be more informed by the tails than the median. We find that differences in the coefficients explain about 32% of the gap in the tails, but about 42% of the gap around the median. AP and BP households may use their endowments of characteristics differently across the distribution to achieve a level of nutrition. Our re-weighting and specification errors are not statistically significant at the 1 percent level for all quantiles. Below, we present the results from the detailed decomposition of the composition and structure effects into the factors that contribute to each across the distribution of nutrition.

4.2.1 Detailed decomposition of the composition effect

Figure 5 shows the results of the detailed decomposition of the composition and structure effects in panels (a) and (b), respectively. We combine the components of categorical variables as in Firpo et al. (2018) to better illustrate the relationships of categorical variables and HEI.

The difference in the amount of FAH-EPP explains about 11.9% of the gap in nutrition at Q5 (about 0.13 HEI points), however this decreases quickly moving across the distribution: at Q10, it explains 6.9% and at Q35, it explains 3.0%. At Q75, it only explains about 1.3% although that increases to 2.2% at Q95 (about 0.07 HEI points). Food-assistance programs which can increase FAH-EPP for BP may help ameliorate the gap in nutrition at the lower tail of the distribution, however, as noted above, even complete amelioration of the gap at the lower tail would not result in a meaningful increase in nutrition.

The difference in education explains 25.7% of the composition effect at Q5 and steadily increases across the distribution to 69.1% at Q95. This is large and amounts to about 17.3% of the total difference at Q5 and 45.4% at Q95. As above, this may reflect other unobservables correlated with education. For example, other authors have found that education may proxy for cognitive ability, health literacy, nutrition knowledge, or health behaviors (Aggarwal et al., 2016; Brunello et al., 2016). The finding that there is an education gradient in nutrition contributes to the wide body of literature documenting and investigating the education gradient in health (see e.g., Lynch and von Hippel 2016; Conti et al. 2010; Cutler and Lleras-Muney 2010).

Differences in employment contribute about 30% of the composition effect in the lower tail, 13.2% at the median, and less than 10% at the upper tail. Increasing opportunities for full-time employment among BP may help reduce the gap. Differences in occupation explains less: about 10% at the lower tail, 11.4% at the median, and less than 5% at the upper tail and are driven mostly by differences in white collar jobs.

We find negative estimates for most of the distribution for the contribution of marital status and children, meaning that the differences in these characteristics are associated with an amelioration of the gap in nutrition. Differences in the ages and presence of children mitigate the gap by about 10% at the lower tail, 5.6% at the median, and a little more than 2.0% at the upper tail. Children younger than 6 particularly ameliorate the gap: if BP households were endowed with the same distribution of children younger than 6 as AP, their nutrition would decrease by 0.05 points at Q10 and about twice that at the median and upper tail. It may be that food assistance programs aimed at households with young children (e.g., the Special Supplemental Nutrition Program for Women, Infants, and Children) are particularly effective at increasing nutrition or diet quality as found by Weinfield et al. (2020). The difference in the racial composition of the groups contributes only slightly contributes to the gap across the entire distribution, primarily driven by differences in the amount of households that identify as Black.

4.2.2 Detailed decomposition of the coefficient effect

The difference in the marginal change in HEI associated with an additional dollar of FAH-EPP changes across the distribution: at the lower tail it is positive (about 0.56) and at the upper tail negative (about -2.51). For much of the middle of the distribution it is positive, but 95% confidence intervals include zero. For most of the distribution of HEI, we find some evidence to support that AP are able to use a dollar of FAH-EPP to more effectively (in terms of nutrition returns) than BP. However, it is interesting to note that at Q90 and Q95 this is reversed and BP more effectively use FAH-EPP to achieve nutrition. We believe this is further evidence that researchers should focus on how BP at the upper tail of the distribution are able to achieve higher HEIs.

While we find that nutrition returns to primary school are greater for AP; the returns to some high school are only statistically greater for AP at the median and above. We do not find statistically significant evidence of differences in returns to other levels of education, except post-graduate degrees. For households at Q85 to Q95, we find that BP are more effectively able to use their post-graduate degree to achieve nutrition than AP, mitigating the nutrition gap by about 0.27 to 0.73 HEI points. Are highly educated BP different from less educated BP in important ways? Comparisons of group compositions at the mean do not yield great insights: a greater percentage is employed full-time, married or single, have children under 6, and are white-collar professionals; a smaller percentage is not employed (i.e., student or homemaker), has no occupation, and have children between 6 & 13. These households do not have a meaningful difference in age or household size. How these BP households are able to achieve such high nutrition returns to post-graduate education appears to be a fruitful area for future research.

We do not find evidence of a difference in nutrition returns to race in general. AP that identify as Asian have statistically greater returns to nutrition than BP at the upper half of the distribution, but the contribution is small. We do not find evidence supporting a difference in returns to identifying as black. Returns to white- and blue-collar occupations are different for households in the upper half of the distribution and the difference contributes about 0.74 HEI points at Q55 to about 2.11 at Q95. We do not find evidence that the differences in returns to employment contribute to the gap in nutrition.There is some evidence across the distribution that the returns to children, particularly those between 6 &13, contribute to the gap across the distribution. We find that BP have greater returns to not being single than AP and that this has a slight mitigating association with the gap in nutrition at the upper and lower tails. For example, returns to marriage are about 1.40 to 2.03 HEI points greater for BP at Q85 - Q95; getting divorced also is associated with slightly higher returns for BP at those quantiles.

5 Conclusions

We investigate how differences in endowments of and returns to food expenditures, demographics, and household factors across AP and BP contribute to the gap in nutritional quality using a recently developed decomposition method designed for statistics beyond the mean (Firpo et al., 2018) and new data sets linking household purchase scanner data with nutrients (Carlson et al., 2019). We find that the gap in nutritional quality between households above and below 130% of the poverty threshold is not consistent across the distribution. At the lower end of the distribution of nutrition, where households are most at risk for diseases related to poor nutrition, the gap is less than 1.5 HEI points; the reduction of which is not expected to confer meaningful benefits. Above the median, the gap is substantially larger; however the maximum difference in

HEI scores between AP and BP is only about 4 HEI points. In contrast, the difference in HEI scores between households at the 90th and 10th quantiles is about 30 points for each household type. Two consequences of this finding are: (1) alleviating poverty may not meaningfully raise the nutrition of households most at risk, and (2) policies that seek to raise nutrition regardless of poverty status may be needed.

It has been suggested that there is a cycle of poor nutrition and chronic diseases among households in poverty, in which households with low income and poor diet may be less able to afford medications or treatment as well as a healthy diet, thus worsening their conditions with even poorer nutrition (Seligman et al., 2010). Under this paradigm, even a small initial difference in nutrition could translate to a much larger disparity in disease incidence. The results of our static model, which does not capture dynamics, do not seem consistent with this hypothesis: if nutrition-induced ill BP purchased an ever-decreasing-in-nutrition basket of food, we would expect the difference in HEI scores to be largest below the median. We find the opposite.

We find that differences in the composition of each group contributes about 68% to the gap in HEI scores at both tails and about 57% around the median. Differences in food expenditures contribute less than 5% for most of the distribution and, referring to food-at-home purchases alone, it is not clear that healthier diets require a greater expenditure. Our analysis is limited to food-at-home purchases and does not include purchases from fast-food outlets, restaurants, or schools, for example, which may inflate HEIs for households that consume a significant portion of their diet from food-away-from-home. In contrast, differences in education and employment contribute about 38% at Q5 to 50% at Q95. We also find that BP above Q80 experience greater returns, in terms of nutrition, to post-graduate education than AP and this slightly mitigates the gap in nutrition at the upper tail. How these households are able to achieve a greater return to HEI from education appears to be a fruitful area for further research.

Decomposition methods, in general, do not seek to recover or identify underlying behavioral mechanisms and we investigate associations with nutrition and do not identify causal relationships. However, by indicating which factors are quantitatively important, and which may not be, this analysis provides useful indications of particular explanations that should be explored further. The FFL decomposition also implicitly rules out general equilibrium effects. Future research should also consider the component scores when investigating differences in HEIs among at-risk populations. We also suggest that comparisons along the BP distribution may reveal useful mechanisms to increase nutrition. Diet- and nutrition-related disease disparities across poverty status are often assumed to be due to differences in nutrition. The results from our analysis provide reasons to believe that this assumption may not be justified and that researchers should focus on identifying other differences– for example, access to timely health care– that may have a more meaningful contribution to the unequal incidence of nutrition-related illnesses.

References

Aggarwal, A., Rehm, C. D., Monsivais, P., and Drewnowski, A. (2016). Importance of taste, nutrition, cost and convenience in relation to diet quality: Evidence of nutrition resilience among us adults using national health and nutrition examination survey (nhanes) 2007–2010. *Preventive medicine*, 90:184–192.

Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., and Schnell, M. (2019). Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics*, 134(4):1793–1844.

Attanasio, O., Blundell, R., Conti, G., and Mason, G. (2020). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, 191:104171.

Ball, K., Crawford, D., and Mishra, G. (2006). Socio-economic inequalities in women's fruit and vegetable intakes: a multilevel study of individual, social and environmental mediators. *Public health nutrition*, 9(5):623–630.

Bann, D., Fitzsimons, E., and Johnson, W. (2020). Determinants of the population health distribution: an illustration examining body mass index. *International journal of epidemiology*, 49(3):731–737.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, 94(4):991–1013.

Bilger, M., Kruger, E. J., and Finkelstein, E. A. (2017). Measuring socioeconomic inequality in obesity: looking beyond the obesity threshold. *Health economics*, 26(8):1052–1066.

Bonanno, A., Bimbo, F., Cleary, R., and Castellari, E. (2018). Food labels and adult bmi in italy–an unconditional quantile regression approach. *Food Policy*, 74:199–211.

Broda, C., Leibtag, E., and Weinstein, D. E. (2009). The role of prices in measuring the poor's living standards. *The Journal of Economic Perspectives*, 23(2):77–97.

Brunello, G., Fort, M., Schneeweis, N., and Winter-Ebmer, R. (2016). The causal effect of education on health: What is the role of health behaviors? *Health economics*, 25(3):314–336.

Carlson, A. C., Page, E. T., Palmer, T., Zimmerman, C. E. T., and Hermansen, S. (2019). Linking usda nutrition databases to iri household-based and store-based scanner data. *U.S. Department of Agriculture, Economic Research Service*, March(TB-1952).

Carrieri, V. and Jones, A. M. (2017). The income–health relationship 'beyond the mean': New evidence from biomarkers. *Health economics*, 26(7):937–956.

Carrieri, V. and Jones, A. M. (2018). Inequality of opportunity in health: A decomposition-based approach. *Health economics*, 27(12):1981–1995.

Churchill, S. A., Munyanyi, M. E., Prakash, K., and Smyth, R. (2020). Locus of control and the gender gap in mental health. *Journal of Economic Behavior & Organization*, 178:740–758.

Conti, G., Heckman, J., and Urzua, S. (2010). The education-health gradient. *American Economic Review*, 100(2):234–38.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of health economics*, 29(1):1–28.

Davillas, A. and Jones, A. M. (2020). Ex ante inequality of opportunity in health, decomposition and distributional analysis of biomarkers. *Journal of Health Economics*, 69:102251.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Duclos, J.-Y. and Échevin, D. (2011). Health and income: A robust comparison of canada and the us. *Journal of Health Economics*, 30(2):293–302.

Firpo, S., Fortin, N., and Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, 6(2):28.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In *Handbook of labor economics*, volume 4, pages 1–102. Elsevier.

Guenther, P. M., Casavale, K. O., Reedy, J., Kirkpatrick, S. I., Hiza, H. A., Kuczynski, K. J., Kahle, L. L., and Krebs-Smith, S. M. (2013). Update of the healthy eating index: Hei-2010. *Journal of the Academy of Nutrition and Dietetics*, 113(4):569–580.

Guenther, P. M., Kirkpatrick, S. I., Reedy, J., Krebs-Smith, S. M., Buckman, D. W., Dodd, K. W., Casavale, K. O., and Carroll, R. J. (2014). The healthy eating index-2010 is a valid and reliable measure of diet quality according to the 2010 dietary guidelines for americans. *The Journal of nutrition*, 144(3):399–407.

Harmon, B. E., Boushey, C. J., Shvetsov, Y. B., Ettienne, R., Reedy, J., Wilkens, L. R., Le Marchand, L., Henderson, B. E., and Kolonel, L. N. (2015). Associations of key diet-quality indexes with mortality in the multiethnic cohort: the dietary patterns methods project. *The American journal of clinical nutrition*, 101(3):587–597.

Haughton, J. and Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. World Bank Publications.

Heckley, G., Gerdtham, U.-G., and Kjellsson, G. (2016). A general method for decomposing the causes of socioeconomic inequality in health. *Journal of health economics*, 48:89–106.

Hiza, H. A., Casavale, K. O., Guenther, P. M., and Davis, C. A. (2013). Diet quality of americans differs by age, sex, race/ethnicity, income, and education level. *Journal of the Academy of Nutrition and Dietetics*, 113(2):297–306.

Hoenig, J. M. and Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55(1):19–24.

Jones, A. M. and Nicolás, A. L. (2004). Measurement and explanation of socioeconomic inequality in health with longitudinal data. *Health economics*, 13(10):1015–1030.

Kant, A. K. and Graubard, B. I. (2007). Secular trends in the association of socio-economic position with self-reported dietary attributes and biomarkers in the us population: National Health and Nutrition Examination Survey (NHANES) 1971–1975 to nhanes 1999–2002. *Public health nutrition*, 10(2):158–167.

Kirkpatrick, S. I., Reedy, J., Krebs-Smith, S. M., Pannucci, T. E., Subar, A. F., Wilson, M. M., Lerman, J. L., and Tooze, J. A. (2018). Applications of the healthy eating index for surveillance, epidemiology, and intervention research: considerations and caveats. *Journal of the Academy of Nutrition and Dietetics*, 118(9):1603–1621.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Leung, C. W., Ding, E. L., Catalano, P. J., Villamor, E., Rimm, E. B., and Willett, W. C. (2012). Dietary intake and dietary quality of low-income adults in the supplemental nutrition assistance program. *The American journal of clinical nutrition*, pages ajcn–040014.

Leung, C. W. and Tester, J. M. (2019). The association between food insecurity and diet quality varies by race/ethnicity: An analysis of national health and nutrition examination survey 2011-2014 results. *Journal of the Academy of Nutrition and Dietetics*, 119(10):1676–1686.

Lynch, J. L. and von Hippel, P. T. (2016). An education gradient in health, a health gradient in education, or a confounded gradient in both? *Social Science & Medicine*, 154:18–27.

Makdissi, P. and Yazbeck, M. (2014). Measuring socioeconomic health inequalities in presence of multiple categorical information. *Journal of Health Economics*, 34:84–95.

Mancino, L., Guthrie, J., Ver Ploeg, M., and Lin, B.-H. (2018). Nutritional quality of foods acquired by americans: findings from usda's national household food acquisition and purchase survey. Technical report.

Mozaffarian, D., Fleischhacker, S., and Andrés, J. R. (2021). Prioritizing nutrition security in the us. *JAMA*, 325(16):1605–1606.

Muth, M. K., Sweitzer, M., Brown, D., Capogrossi, K., Karns, S. A., Levin, D., Okrent, A., Siegel, P., and Zhen, C. (2016). Understanding iri household-based and store-based scanner data. Technical report.

National Research Council and the Institute of Medicine (2013). *Supplemental nutrition assistance program: examining the evidence to define benefit adequacy*. National Academies Press.

Onvani, S., Haghighatdoost, F., Surkan, P., Larijani, B., and Azadbakht, L. (2017). Adherence to the healthy eating index and alternative healthy eating index dietary patterns and mortality from all causes, cardiovascular disease and cancer: a meta-analysis of observational studies. *Journal of Human Nutrition and Dietetics*, 30(2):216–226.

Pappas, G., Queen, S., Hadden, W., and Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the united states, 1960 and 1986. *New England journal of medicine*, 329(2):103–109.

Pradhan, M., Sahn, D. E., and Younger, S. D. (2003). Decomposing world health inequality. *Journal of health economics*, 22(2):271–293.

Provencher, A. and Carlton, A. (2018). The poverty experience of lone mothers and their children. *Applied Economics Letters*, 25(6):401–404.

Rehm, C. D., Monsivais, P., and Drewnowski, A. (2015). Relation between diet cost and healthy eating index 2010 scores among adults in the united states 2007–2010. *Preventive medicine*, 73:70–75.

Rehm, C. D., Peñalvo, J. L., Afshin, A., and Mozaffarian, D. (2016). Dietary intake among us adults, 1999-2012. *Jama*, 315(23):2542–2553.

Rycroft, R. S. (2017). *The economics of inequality, discrimination, poverty, and mobility*. Routledge.

Schiller, B. R. (2004). *Economics of Poverty and Discrimination*. Number 9th. Pearson Prentice Hall.

Seligman, H. K., Schillinger, D., et al. (2010). Hunger and socioeconomic disparities in chronic disease. *New England Journal of Medicine*, 363(1):6–9.

Silbersdorff, A., Lynch, J., Klasen, S., and Kneib, T. (2018). Reconsidering the income-health relationship using distributional regression. *Health economics*, 27(7):1074–1088.

Singleton, C. R., Young, S. K., Kessee, N., Springfield, S. E., and Sen, B. P. (2020). Examining disparities in diet quality between snap participants and non-participants using oaxaca-blinder decomposition analysis. *Preventive Medicine Reports*, 19:101134.

Smith, T. A., Valizadeh, P., Lin, B.-H., and Coats, E. (2019). What is driving increases in dietary quality in the united states? *Food Policy*, page 101720.

Sweitzer, M., Brown, D., Karns, S., Muth, M. K., Siegel, P., and Zhen, C. (2017). Food-at-home expenditures: Comparing commercial household scanner data from iri and government survey data. *Washington, DC: Economic Research Service, United States Department of Agriculture.*

Tiffin, R. and Arnoult, M. (2010). The demand for a healthy diet: estimating the almost ideal demand system with infrequency of purchase. *European Review of Agricultural Economics*, 37(4):501–521.

U.S. Department of Agriculture and U.S. Department of Health and Human Services (2010). *Dietary Guidelines for Americans 2010*.

Variyam, J. N., Blaylock, J., and Smallwood, D. (2002). Characterizing the distribution of macronutrient intake among us adults: a quantile regression approach. *American Journal of Agricultural Economics*, 84(2):454–466.

Weinfield, N. S., Borger, C., Au, L. E., Whaley, S. E., Berman, D., and Ritchie, L. D. (2020). Longer participation in wic is associated with better diet quality in 24-month-old children. *Journal of the Academy of Nutrition and Dietetics*, 120(6):963–971.

Zhen, C., Muth, M., Okrent, A., Karns, S., Brown, D., and Siegel, P. (2019). Do differences in reported expenditures between household scanner data and expenditure surveys matter in health policy research? *Health economics*, 28(6):782–800.

Table 1: Summary statistics, estimated using survey weights

	Above 130%		Below 130%	
Variable	Mean	S.D.	Mean	S.D.
HEI-2010	44.4693	0.0212	41.6199	0.0582
FAH expenditures per person	106.1009	0.1144	100.5932	0.3380
Education				
Primary school	0.0013	0.0001	0.0076	0.0004
High school diploma	0.1192	0.0005	0.2667	0.0020
Some high school	0.0060	0.0001	0.0380	0.0009
Some college	0.2571	0.0007	0.3671	0.0023
College diploma	0.4065	0.0008	0.2621	0.0021
Post graduate	0.2100	0.0007	0.0585	0.0011
Race & ethnicity				
Hispanic	0.1121	0.0007	0.1141	0.0020
White	0.7858	0.0007	0.7553	0.0023
Black	0.1076	0.0005	0.1408	0.0018
Asian	0.0432	0.0004	0.0218	0.0008
Other race	0.0634	0.0005	0.0821	0.0015
Occupation				
White collar	0.6313	0.0008	0.2697	0.0022
Blue collar	0.1135	0.0005	0.1126	0.0016
No collar	0.2552	0.0006	0.6177	0.0024
Ages & presence of children				
Children present under 6	0.1261	0.0007	0.1502	0.0021
Children present between 6 & 13	0.1580	0.0007	0.2080	0.0022
Children present between 13 & 18	0.1399	0.0006	0.1786	0.0019
No children	0.6969	0.0009	0.6532	0.0024
Marital status				
Married	0.6729	0.0008	0.4085	0.0024
Widowed	0.0603	0.0003	0.1036	0.0012
Divorced	0.1169	0.0004	0.2309	0.0018
Single	0.1499	0.0007	0.2570	0.0023
Employment				
Employed part-time	0.0904	0.0004	0.2083	0.0021
Employed full-time	0.7075	0.0007	0.2716	0.0022
Not employed	0.2021	0.0005	0.5201	0.0024
Other characteristics				
Household size	2.5707	0.0026	2.5843	0.0088
Age	53.2478	0.0283	51.9428	0.0855

NB: State- and month-fixed effects are omitted here for brevity.

Table 2: Estimated differences in dietary quality for households above and below 130% of the poverty threshold for selected quantiles

	FFL Decompositions							Inequality Measures		
Variable	Q05	Q10	Q25	Q50	Q75	Q90	Q95	Q90-Q10	Q50-Q10	Q90-Q50
Above 130% of the poverty threshold	25.4342 ***	28.9891 ***	34.7285 ***	43.2801 ***	53.4302 ***	62.3687 ***	67.3690 ***	33.3796 ***	14.2911 ***	19.0886 ***
	(0.0159)	(0.0211)	(0.0149)	(0.0217)	(0.0130)	(0.0396)	(0.0340)			
Below 130% of the poverty threshold	24.3578 ***	27.6102 ***	32.8840 ***	40.0483 ***	49.5496 ***	58.5404 ***	63.7724 ***	30.9302 ***	12.4382 ***	18.4921 ***
	(0.0937)	(0.0463)	(0.0351)	(0.0650)	(0.0638)	(0.0810)	(0.0629)			
Total difference	1.0764 ***	1.3789 ***	1.8445 ***	3.2318 ***	3.8806 ***	3.8283 ***	3.5966 ***	2.4494 ***	1.8529 ***	0.5965 ***
	(0.0962)	(0.0598)	(0.0414)	(0.0716)	(0.0693)	(0.1034)	(0.0701)			
Compositional effect	0.7244 ***	0.9199 ***	1.0388 ***	1.8523 ***	2.1822 ***	2.6422 ***	2.3606 ***	1.7222 ***	0.9324 ***	0.7899 ***
	(0.1504)	(0.1419)	(0.1470)	(0.1418)	(0.2872)	(0.1218)	(0.1402)			
Structural effect	0.3520 ***	0.4590 ***	0.8057 ***	1.3795 ***	1.6984 ***	1.1861 ***	1.2360 ***	0.7271 ***	0.9205 ***	-0.1934 ***
	(0.1353)	(0.1440)	(0.1334)	(0.1157)	(0.2960)	(0.1197)	(0.1528)			
Reweighting error	-0.1106	-0.1390	-0.1453	-0.2799	-0.3482	-0.3302	-0.3725			
	(0.2229)	(0.2330)	(0.1911)	(0.2339)	(0.3000)	(0.3268)	(0.4330)			
Specification error	0.0267	0.0661	0.0262	-0.0242	-0.0408	0.1397	0.1239			
	(0.2868)	(0.2560)	(0.2025)	(0.2734)	(0.3643)	(0.4570)	(0.6227)			

NB: *** indicates statistical significance at less than 1%, ** at 5%, and * at 10%.

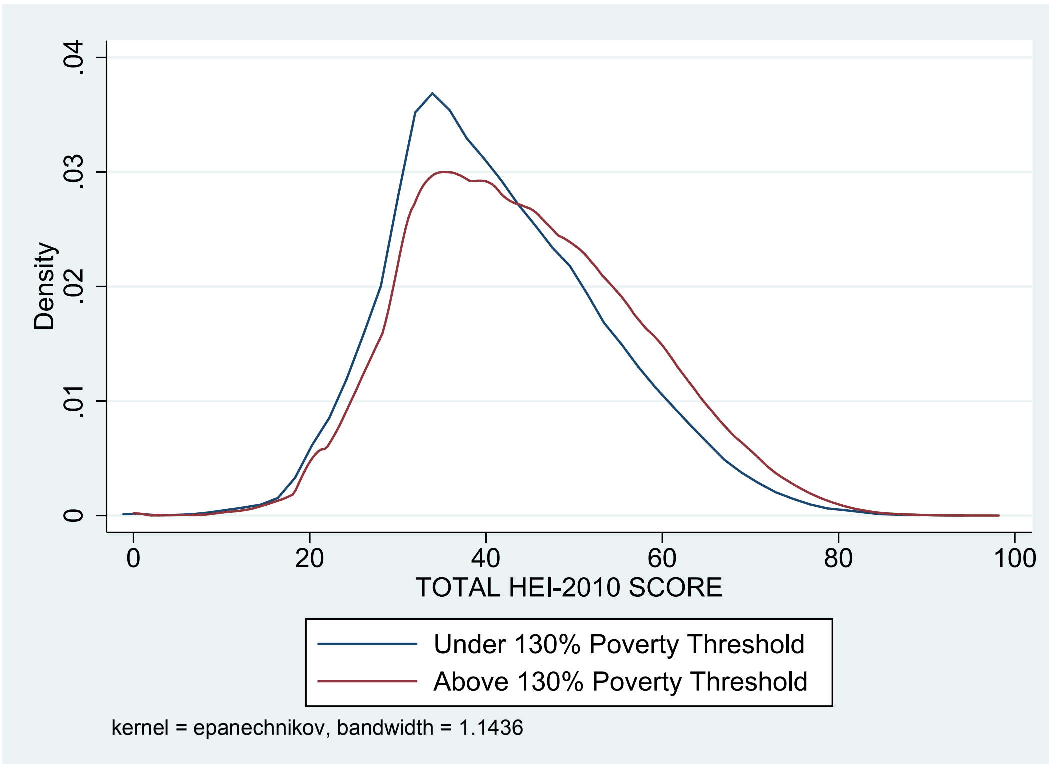


Figure 1: **Kernel densities of HEI for households above and below 130% of the poverty threshold**. This figure shows HEI density functions for both BP and AP households. The HEI distribution of the BP household has a higher mean and greater spread than that of the BP. We do not find evidence of heaping across either distribution, despite the HEI components being top- and bottom-coded. The distributions are quite smooth; for examples of problematic heaping see Firpo et al. (2018).

Figure 2: Estimated coefficients across quantiles of HEI for households above 130% of the poverty threshold

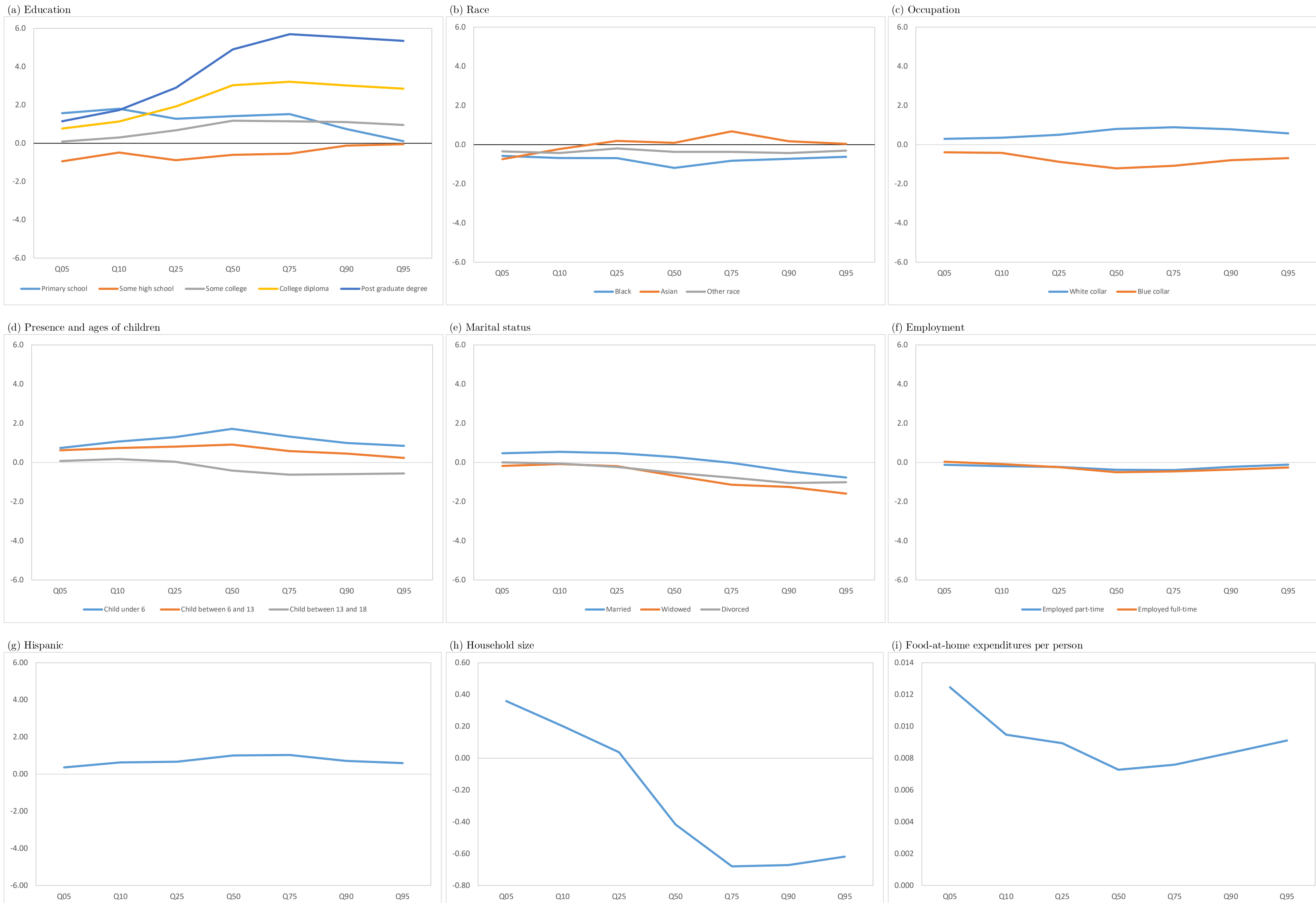


Figure 3: Estimated coefficients across quantiles of HEI for households below 130% of the poverty threshold

(a) Aggregate decomposition (b) Detailed total decomposition

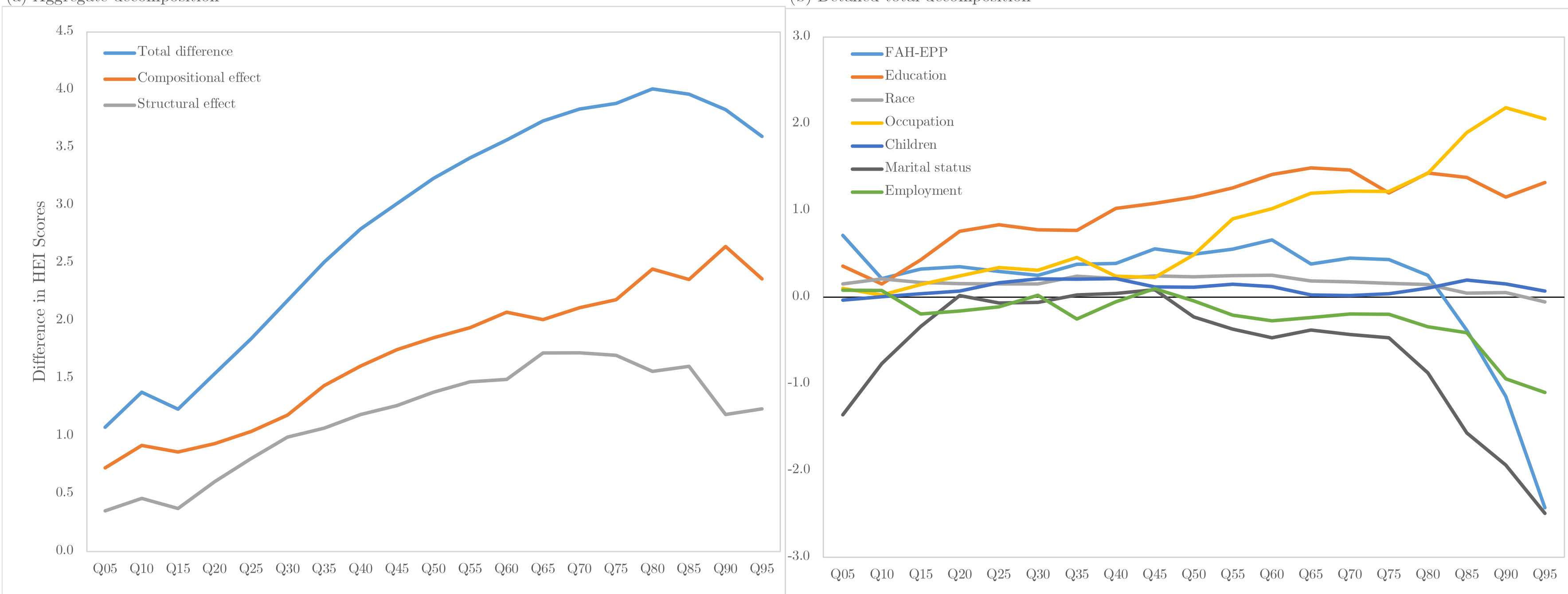


Figure 4: **Differences in HEI scores across poverty levels**. Differences are the result of above 130% of the poverty threshold statistic minus below 130% of the poverty threshold statistic. Differences are small at the lower tail and get larger toward the upper tail, indicating that measures to overcome poverty alone may not increase nutritional qualities for those households most at-risk of diet-related diseases. The composition effect dominates the structural effect, particularly toward the higher values of the HEI distribution, suggesting that a redistribution policy may be most effective to reduce inequality in dietary quality for households with higher than median dietary qualities.

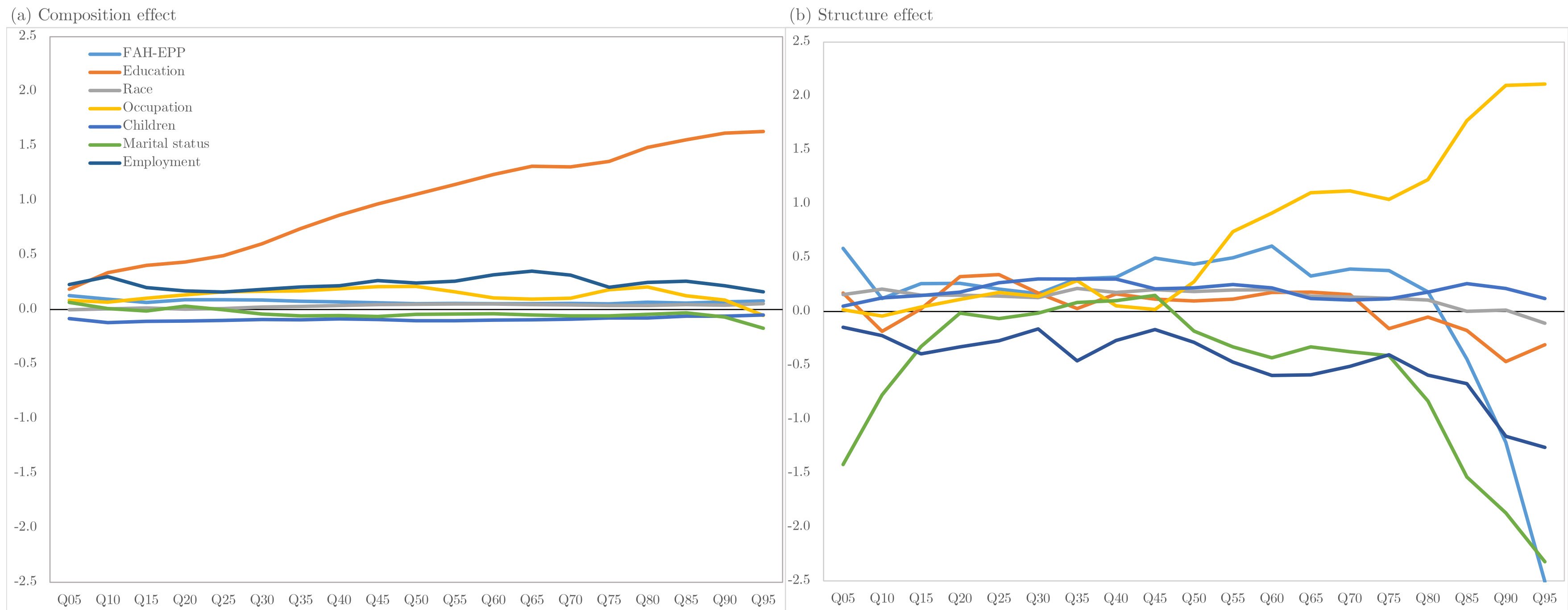
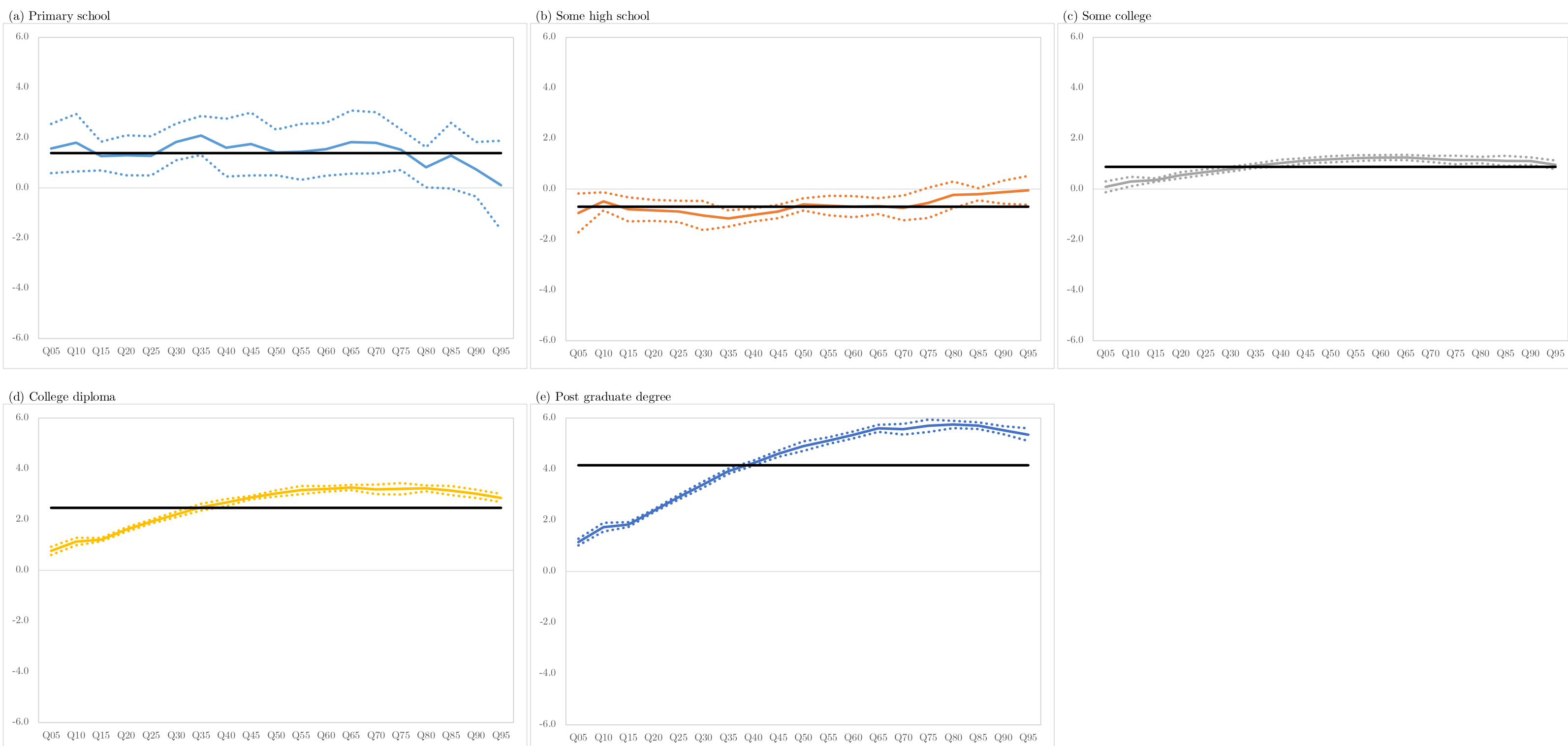


Figure 5: **Decomposition of contributing factors to the composition and structural effects**. Panel (a) shows the possible HEI changes for BP under the counterfactual of endowing BP with the same levels of covariates as AP . Panel (b) shows the possible HEI changes for BP if they were to have the same returns (reweighted for the BP distribution) as AP.

Appendix.

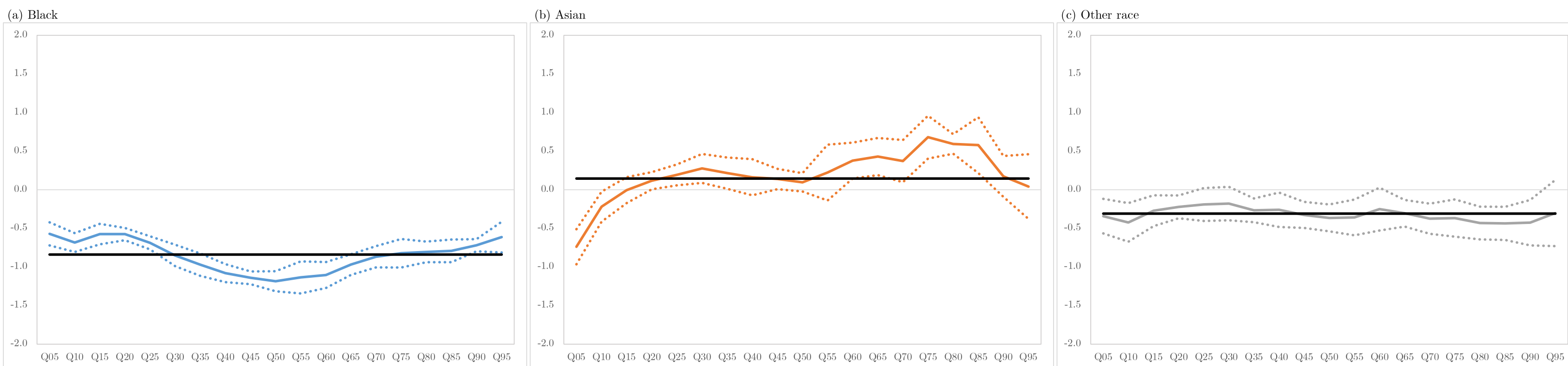
We include figures A1 to A14 showing the confidence intervals of the UQR estimates across quantiles compared to the OLS estimate. These figures can be interpreted as an equivalence test between the estimated UQR parameter at a quantile and the OLS estimate. If the OLS estimate lies outside the 95% confidence intervals of the UQR estimate at a quantile, we reject equivalence between the OLS and UQR estimates at that quantile at the 5% level; see Hoenig and Heisey (2001) for more details on equivalence testing.

Figure A1: Equivalence tests of the UQR and OLS coefficients for households above 130 % of the poverty threshold: education



NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

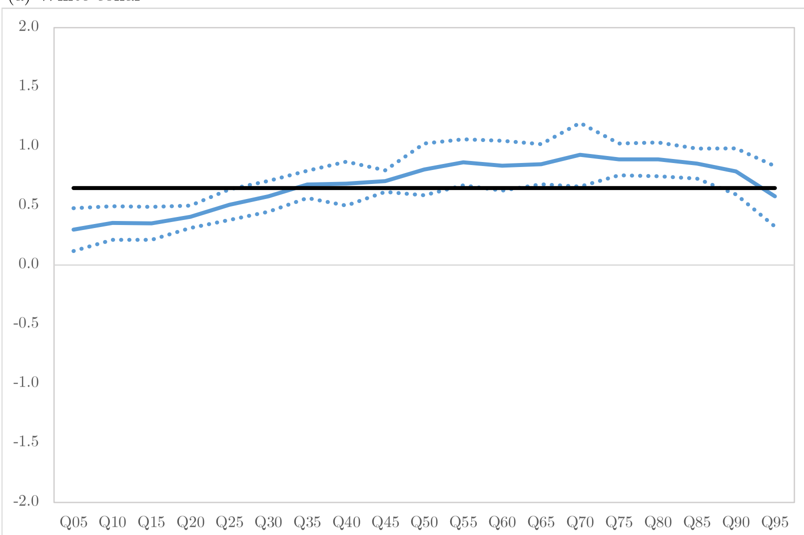
Figure A2: Eqivalence tests of the UQR and OLS coefficients for households above 130% of the poverty threshold: race



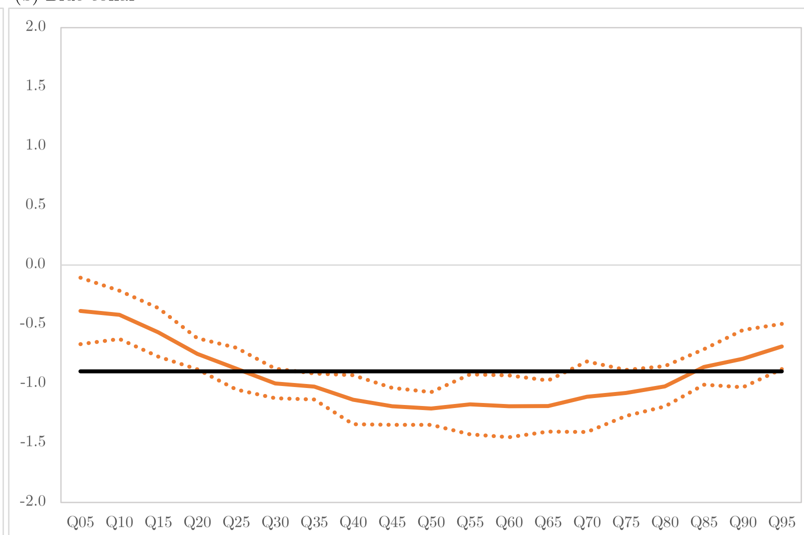
NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A3: Equivalence tests of the UQR and OLS coefficients for households above 130% of the poverty threshold: occupation

(a) White collar



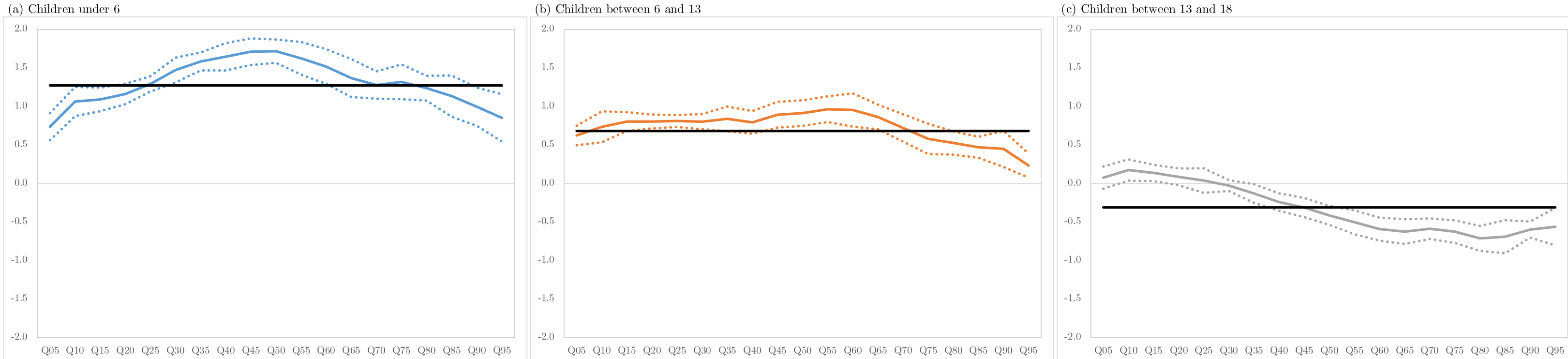
(b) Blue collar



NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

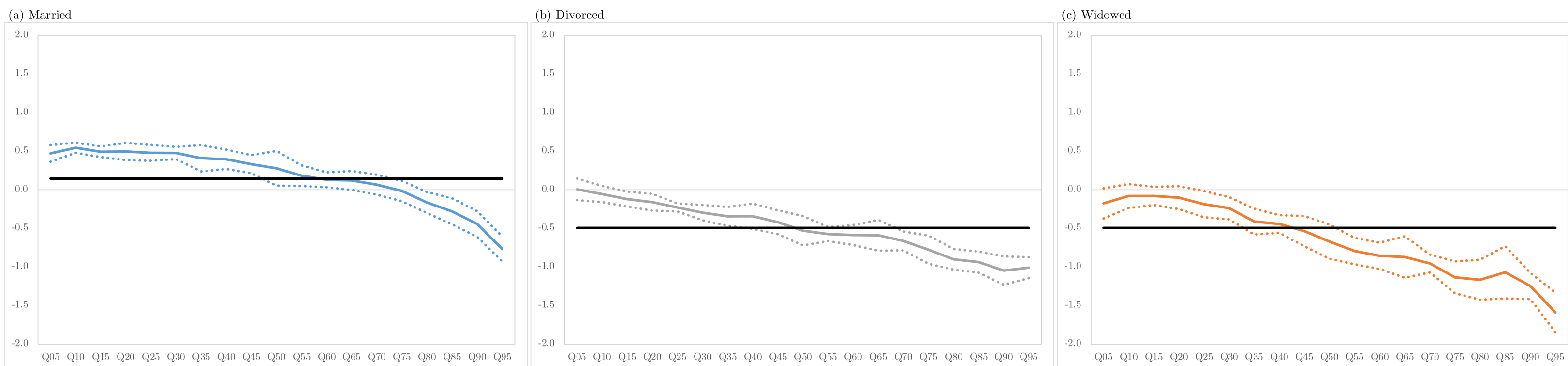
Figure A4: Equivalence tests of the UQR and OLS coefficients for households above 130% of the poverty threshold: ages and presence of children

(a) Children under 6 (b) Children between 6 and 13 (c) Children between 13 and 18



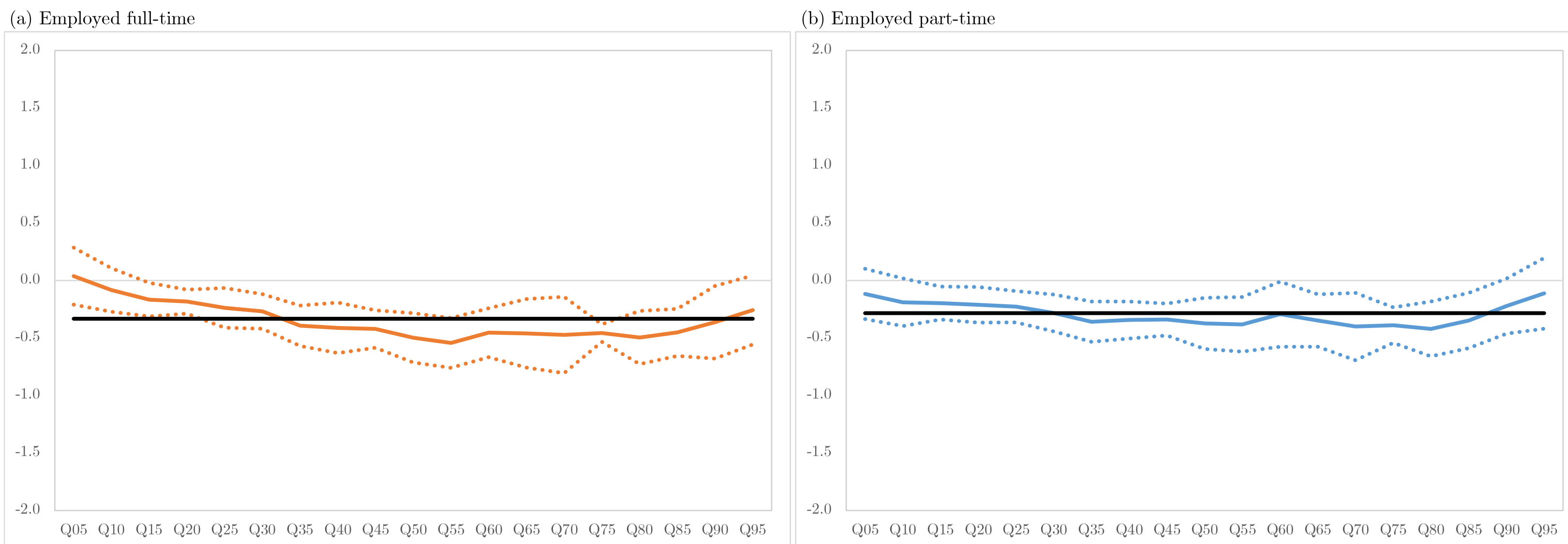
NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A5: Equivalence tests of the UQR and OLS coefficients for households above 130% of the poverty threshold: marital status



NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

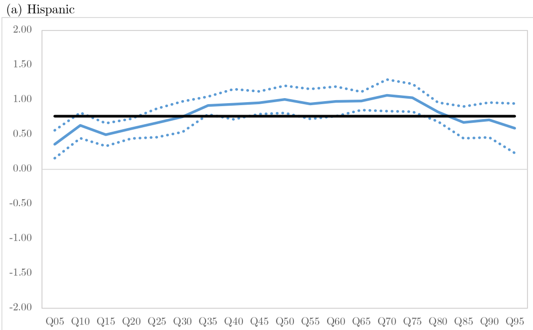
Figure A6: Equivalence tests of the UQR and OLS coefficients for households above 130% of the poverty threshold: employment



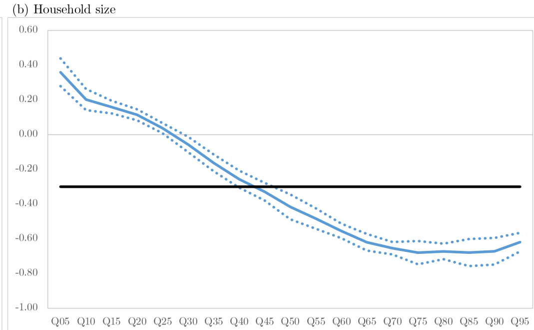
NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A7: Equivalence tests of the UQR and OLS coefficients for households above 130% of the poverty threshold: other

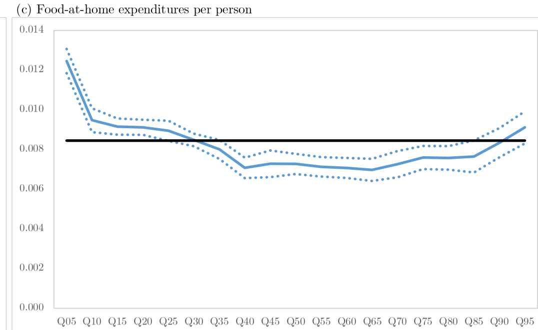
(a) Hispanic



(b) Household size

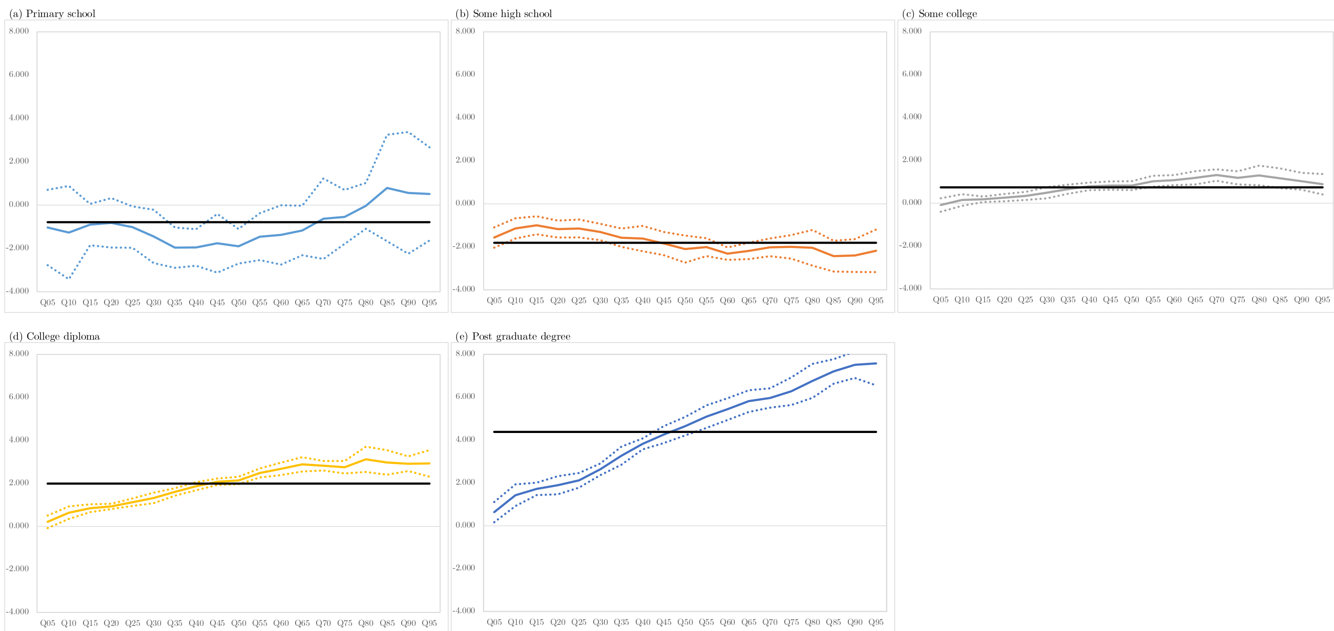


(c) Food-at-home expenditures per person



NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

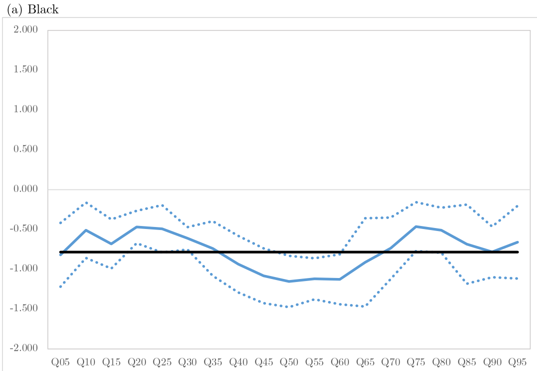
Figure A8: Equivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: education



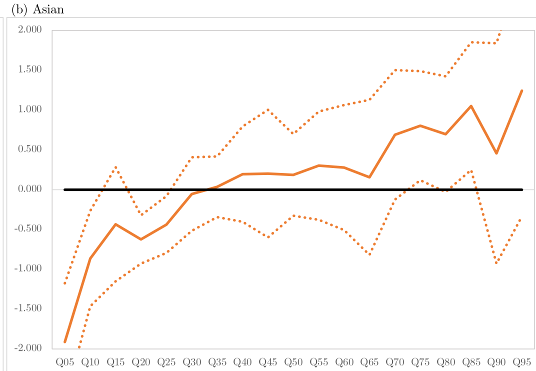
NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A9: Eqivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: race

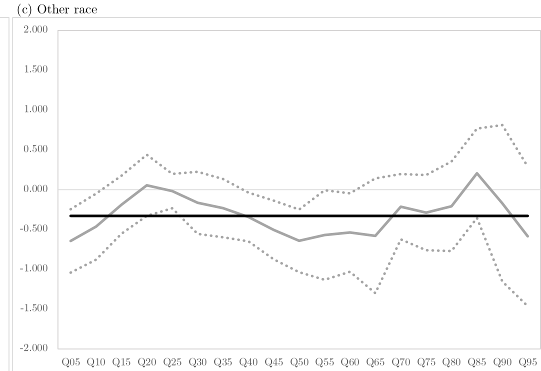
(a) Black



(b) Asian



(c) Other race

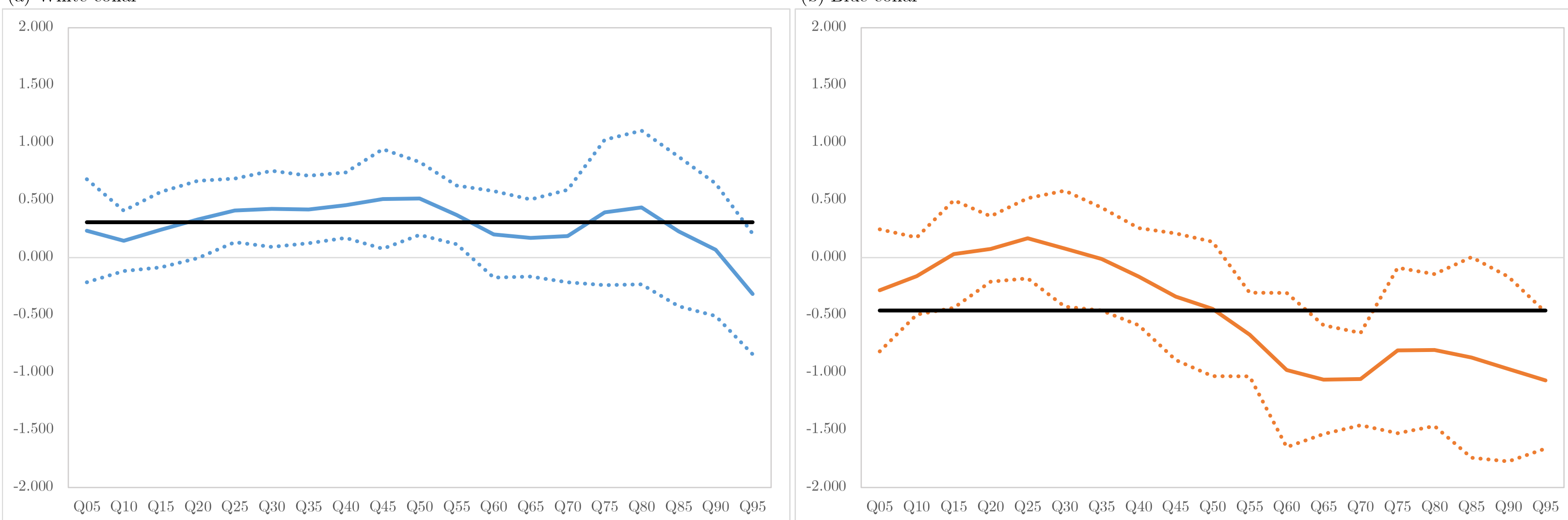


NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A10: Equivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: occupation

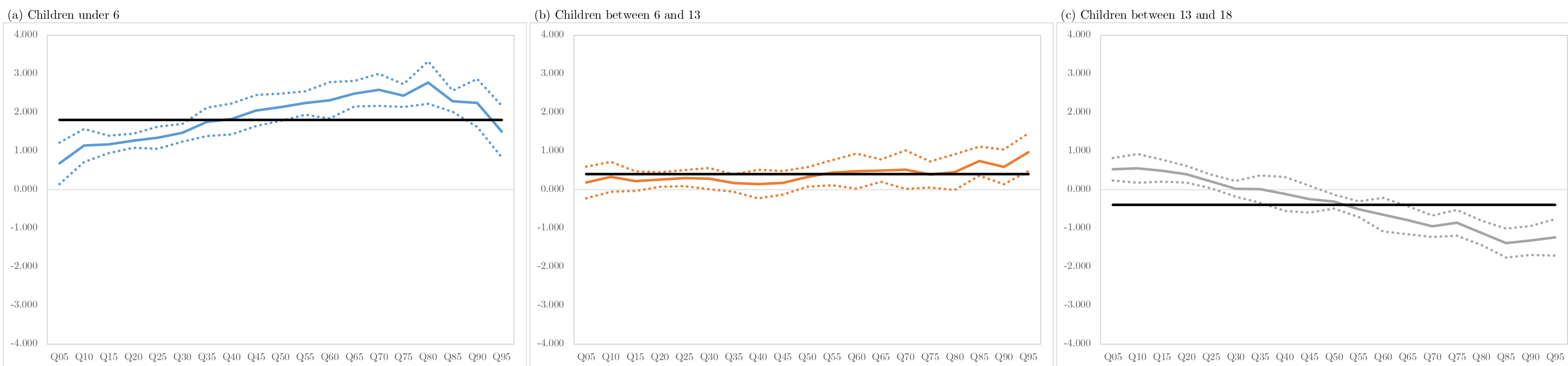
(a) White collar

(b) Blue collar



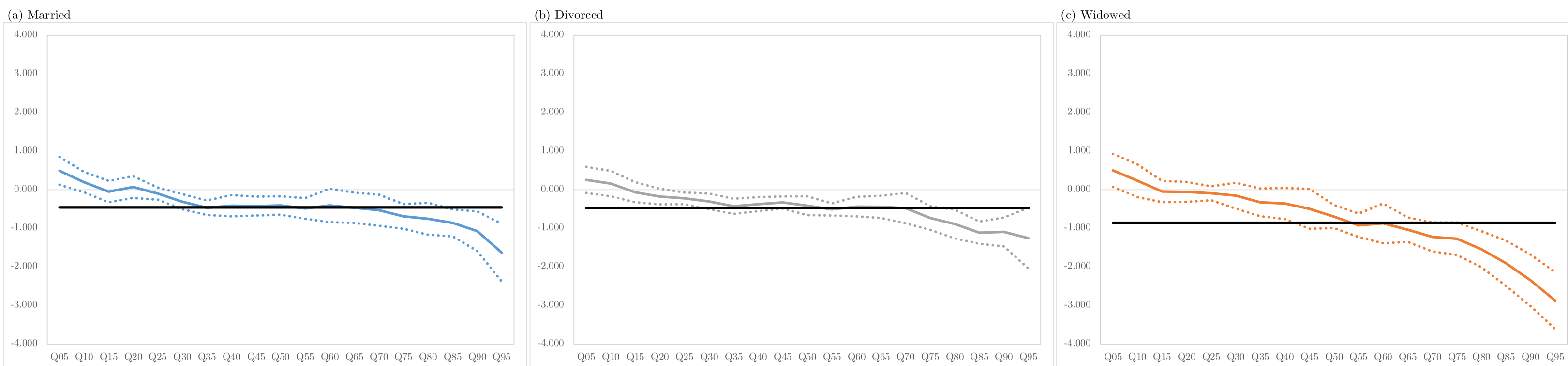
NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A11: Equivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: ages and presence of children



NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A12: Equivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: marital status

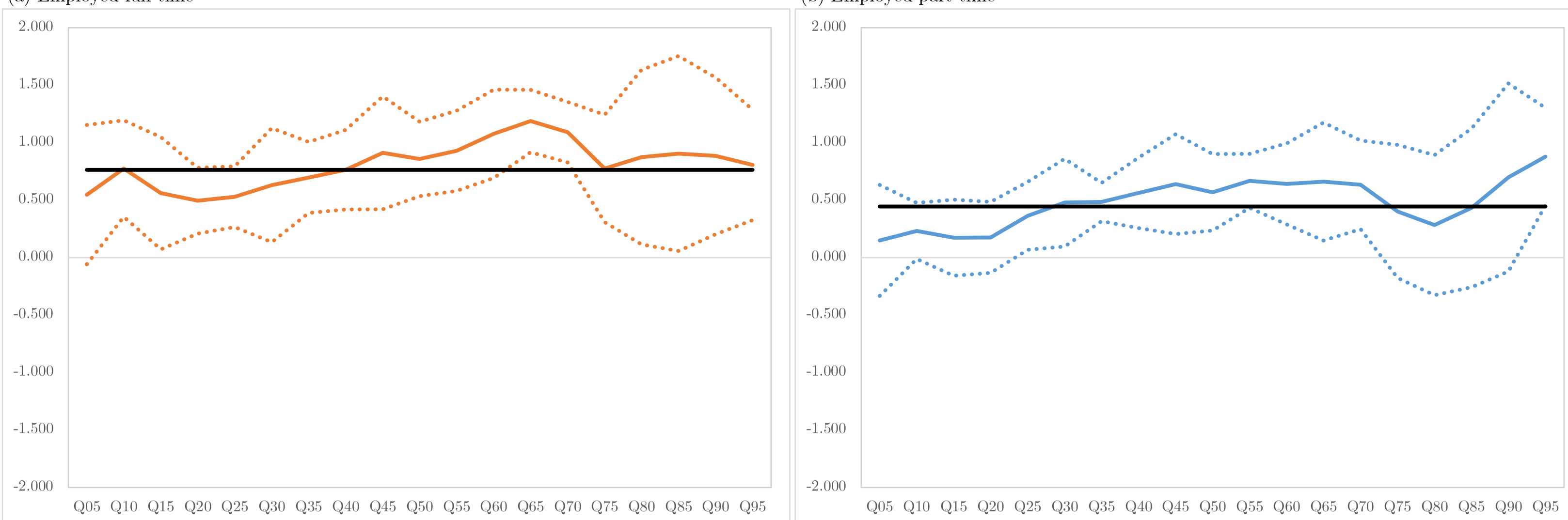


NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A13: Equivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: employment

(a) Employed full-time

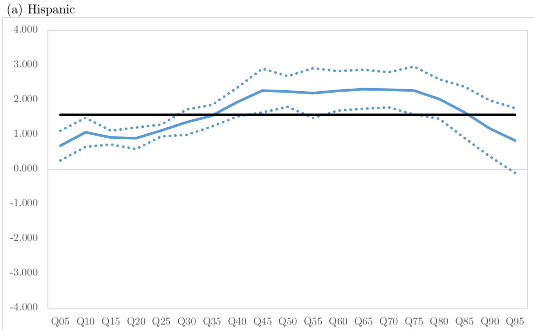
(b) Employed part-time



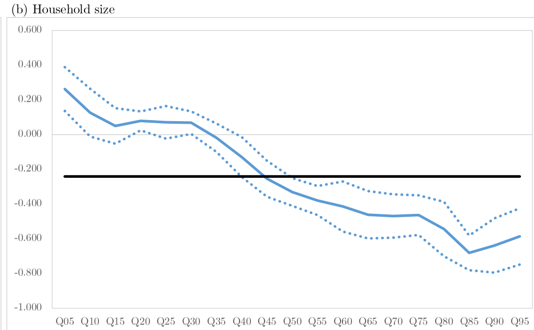
NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.

Figure A14: Equivalence tests of the UQR and OLS coefficients for households below 130% of the poverty threshold: other

(a) Hispanic



(b) Household size



(c) Food-at-home expenditures per person

NB: 95% confidence intervals are given by the dotted lines. The solid black line represents the OLS estimate and the solid colored line represents the UQR estimates.